Is Information in the Brain Represented in Continuous or Discrete Form?

by

James Tee* and Desmond P. Taylor
Communications Research Group,
Department of Electrical and Computer Engineering,
University of Canterbury,
Christchurch, New Zealand.

*Corresponding author: james.tee@canterbury.ac.nz
Department of Electrical & Computer Engineering, University of Canterbury,
Private Bag 4800, Christchurch 8020, New Zealand.

Table of Contents category: Neuroscience – behavioural/systems/cognitive

Keywords: continuous, discrete, brain, representation, quantization, information, computations, communications, coding, probability, intertemporal choice.

Key Points Summary

- The form of information representation in the brain, whether continuous or discrete, is a contentious physiological research question, with an overwhelming majority of contemporary neuroscientists defaulting to the continuous form.
- It is a fact that information (e.g. human vision) must be transmitted over a relatively vast distance (via the optic nerve to the visual cortex) in the presence of noise (e.g. thermal noise, conduction loss, neurobiological imperfections) while preserving its fidelity.
- Shannon's communications theory posits that continuous representation of information in the brain is not plausible under noisy conditions; it has to be in discrete form.
- We test this discrete hypothesis using a computer simulation of neuronal communications, a human behavioral experiment on probability estimation and a human behavioral experiment on intertemporal choices.
- Our extensive results strongly support the discrete hypothesis, implying the need to revisit fundamental assumptions on the brain's physiology.

Abstract

The question of continuous-versus-discrete information representation in the brain is a fundamental yet unresolved physiological question. Historically, most analyses assume a continuous representation without considering the alternative possibility of a discrete representation. Our work explores the plausibility of both representations, and answers the question from a communications engineering perspective. Drawing on the well-established Shannon's communications theory, we posit that information in the brain is represented in a discrete form. Using a computer simulation, we show that information cannot be communicated reliably between neurons using a continuous representation, due to the presence of noise; neural information has to be in a discrete form. In addition, we designed 3 (human) behavioral experiments on probability estimation and analyzed the data using a novel discrete (quantized) model of probability. Under a discrete model of probability, two distinct probabilities (say, 0.57 and 0.58) are treated indifferently. We found that data from all participants were better fit to discrete models than continuous ones. Furthermore, we re-analyzed the data from a published (human) behavioral study on intertemporal choice using a novel discrete (quantized) model of intertemporal choice. Under such a model, two distinct time delays (say, 16 days and 17 days) are treated indifferently. We found corroborating results, showing that data from all participants were better fit to discrete models than continuous ones. In summary, all results reported here support our discrete hypothesis of information representation in the brain, which signifies a major demarcation from the current understanding of the brain's physiology.

Introduction

Music is composed of continuous vibrations of sound waves. These waves can be recorded (or stored) in either continuous (e.g. cassette tape, LP record) or discrete (e.g. CD, MP3) form. This equivalence leads to a fundamental research question in neuroscience, namely, is information in the brain represented in continuous or discrete form? To clarify, there is no doubt that neural signals (electrical waveforms) are continuous-valued (e.g. a typical cell membrane voltage potential can take any value between -40 mV and -70 mV); the question is then whether the information embedded within (or carried by) these neural signals is continuous or discrete. A helpful analogy here is the Morse Code, where information is clearly discrete (i.e. dots or dashes) but the electrical (or radio) signals that carry this discrete information are continuous. While this neural information representation question may appear to be trivial, it remains unresolved. McCulloch and Pitts (1943) theorized that computation in the brain is digital (i.e. discrete), whereas Lashley (1960) suggested that *"the brain is an analogical machine, not digital"*. More recently, VanRullen and Koch (2003) highlighted that a discrete representation of conscious perception (in the brain) is often *"considered but never widely accepted"*. At the same time, a continuous representation *"cannot satisfactorily account for a large body of psychophysical data"*. In the visual working memory literature, this continuous-versus-discrete debate is currently at an impasse – an extensive review of arguments for and against both positions can be found in Luck and Vogel (2013), and Ma et al. (2014). Perhaps the problem is best summed up by Gallistel (2016): *"We do not yet know in what abstract form (analog or digital) the mind stores the basic numerical quantities that give substance to the foundational abstractions, the information acquired from evidence."* [We note here that our use of the term representation encompasses a broad range of information processing activities (or functions) in the brain, including thinking (performing computations, decision-making) and remembering (forming and recollecting memories). Furthermore, in the present work, the terms continuous and analog are treated as equivalent in an engineering sense, as are the terms discrete and digital.]

In the absence of a clear resolution to this debate, a continuous representation (of information in the brain) is almost always the de facto position when considering models and theories arising from experimental data. This is, in part, due to the appeal of the mathematical elegance and analytical convenience of continuous representations (or Real numbers). For example, in Abbott, DePasquale and Memmesheimer (2016), the authors pointed out that neurons communicate with one another *"almost exclusively through discrete action potentials"* – yet, the models described in the rest of the paper employ continuous representation (derivative calculus using Real numbers). Similarly, in Chaudhuri and Fiete (2016), discrete attractor models were lauded for their superior stability and noise tolerance in the brain – yet, the authors defaulted to continuous attractor models as the null hypothesis. A rational question to pose at this point is: why is this continuous-versus-discrete problem important to brain research?

One answer, as outlined by VanRullen and Koch (2003), is that continuous representation *"cannot satisfactorily account for a large body of psychophysical data"*. For example, 1 cent does not typically have much value to most people. However, a person may decide to buy a product if priced at $1.99 – yet, refuse to buy the same product if priced 1 cent higher at $2.00. Such an abrupt (or step) change in the brain's purchasing decision cannot be modeled using a continuous representation despite extensive attempts to do so (Basu, 1997). In contrast, models based on discrete representations are able to accommodate such an abrupt (or step) change. In decision-making literature, Varshney and Varshney (2017) showed that a Bayesian model using quantized (discrete) priors was able to better model racial discrimination behavior. In monetary economics, Khaw et al. (2017) found that a discrete model was a better fit to data from a price-setting experiment than the "optimal" (continuous) models. In mathematical psychology, Sun et al. (2012) applied a Bayes-optimal quantized (discrete) model of perception to datasets of animal vocalizations and human speech, closely mimicking the Weber-Fechner law. In theoretical neuroscience, Varshney et al. (2006) posited that neurons with discrete-valued synaptic states might perform better than neurons with continuous-valued synaptic states. In a recent electrophysiology experiment, Latimer et al. (2015) demonstrated extensively that a discrete (i.e. stepping) model is a better fit to neural (macaque LIP) data than a continuous

(ramping, diffusion-to-bound) model (Gold & Shadlen, 2007). In machine learning, discrete (i.e. quantized or binary) neural networks have been shown to be superior to continuous neural networks in terms of error rate performance (Lin et al., 2016) and computation speed (Hubara et al., 2016).

Another reason for favoring discrete representation, as suggested by Chaudhuri and Fiete (2016), is noise tolerance. There are many sources of noise in the brain (e.g. sensory noise, cellular noise, motor noise) – an extensive review of these noise sources can be found in Faisal et al. (2008). Due to the presence of noise, information that is represented or transported in continuous form inevitably degrades and becomes corrupted in an irrecoverable manner. Analogously, the quality of music that is recorded (or stored) in continuous form (a cassette tape) will slightly degrade due to noise every time it is replayed. On the other hand, the quality of music that is recorded (or stored) in discrete form (CD) is more robustly preserved in the presence of noise. In a wider context, information in the brain must be transmitted over a relatively long distance while preserving its quality. For example, information in human vision originates at the retina, and is conveyed to the LGN via the optic nerve, before arriving at the visual cortex located at the back of the brain – a distance spanning the entire length of the brain. Applying Shannon's communications theory (Shannon, 1948) to communications in the brain, the presence of noise throughout such a transmission chain makes it unfeasible and implausible for the brain to employ continuous representation; noise will always accumulate and corrupt the information, no matter how small the noise is.

In this paper, we approach this discrete-versus-continuous problem from a communications engineering perspective. Given the existence of noise, the requirements for long-range information transmission, and repeated retrievability/retransmissions (memory recall), we posit that information in the brain has to be represented in discrete form, rather than continuous (see Appendix 1). In subsequent sections, we will explore this discrete hypothesis using 3 approaches: 1) a computer simulation of information communications between 2 neurons; 2) a psychophysics (behavioral) experiment on probability estimation; 3) a re-analysis of a published behavioral experiment on intertemporal choice. In each instance, we clearly demonstrate the superior performance of

discrete representation. Overall, our results conclude that the brain must represent (and process) information in discrete form, and must integrate communications aspects directly with computational processes.

Part 1: A Computer Simulation of Information Communications Between 2 Neurons

This section posits that the brain communicates and computes in a discrete manner. While physical signals may be continuous-valued, the communicated or computed information is discrete in nature. We demonstrate this via a set of simulations.

The Computational Theory of Mind (CTM) (Rescorla, 2016) posits that the mind/brain is a computational system, where mental processes such as reasoning, decision-making and problem solving are computations. Marr (1982) further subdivides these mind/brain computations into 3 levels, namely, computational theory (goals of the computation), representation and algorithms (inputs, transformations and outputs), and hardware implementation (neural realizations). There is, however, a major gap in CTM – it does not include a communications component. Computation alone is not useful, because the results must be transmitted to where they will be used. Otherwise, computation is merely an energy-consuming, heat-generating process. For example, suppose that an encrypted file is to be decrypted by a federal intelligence agency in a brute force manner using a supercomputer. Once the supercomputer has successfully cracked the encryption, the resulting file must be transmitted to a useful destination. Otherwise, it remains inaccessible, and therefore, useless. In fact, the encrypted file has to be transmitted to the supercomputer in the first place, before decryption begins. Similarly in the brain, information must be communicated (or transported) before (and after) computation can be performed. For example, the sight of a rattlesnake, when detected by the eye/retina, has to be transmitted to the brain via the optic nerve and LGN before any visual recognition

(computation) ("this object is a rattlesnake") occurs; once the rattlesnake has been identified, the resulting computational output has to be communicated to another part of the brain so that the appropriate cognitive decision ("avoid this danger") can be computed. Communication is thus a necessary prerequisite to computations in the brain, and therefore, must be integrated with computation.

How may communications be integrated with computations? And, how may communications be related to information representation? The answers are also closely related to the presence of noise in the brain (Faisal et al., 2008). First, it is clear that communications in the brain must occur in the presence of noise, which is ubiquitous. Shannon's communications theory (Shannon, 1948) posits that any communications via electrical signals, radio signals, or fiber optic signals will always occur in the presence of noise, and inherently employs a discrete representation of the information. Noise will inevitably degrade the information if continuous representation is employed. Second, given that the communicated information is represented in discrete form, computations using such an input must also be represented in discrete form. Furthermore, the output of a computation must be communicated to a different part of the brain in discrete form. Adrian (1914) first discovered that propagation of neural spikes follows an all or none principle, implying that spike propagation is binary (digital, discrete). This principle is well understood and accepted in neuroscience in terms of its neurobiological implementation (Rieke et al., 1997). Rushton (1961) concluded that the neural signaling of a typical human myelinated nerve fiber spanning, say, between a finger and the spinal cord cannot employ a continuous representation due to the presence of noise. Despite these seminal works, computational models based on continuous representation dominate present-day neuroscience literature – for example, continuous attractor networks (Eliasmith, 2005; Wang, 2009). The computer simulation work of this section is focused on the neurobiological mechanism of information communications between 2 neurons to demonstrate the distorting and degrading effects of stochastic noise in the brain, particularly when a continuous representation is employed.

METHODS

Rushton (1961) examined the transmission properties of a single myelinated axon with the myelin sheathing delimited at regular intervals by Nodes of Ranvier (NoR). He recognized that Nodes of Ranvier (NoR) serve as boosting (or relay) stations, repeatedly restoring the action potential signal to its initial level (which otherwise would be severely attenuated) during propagation along the axon. Our approach is an adaptation of Rushton (1961). In order to simulate communications or information transmission between 2 neurons connected by a myelinated axon, we employ the communications systems model defined by Shannon (1948) (Figure 1).

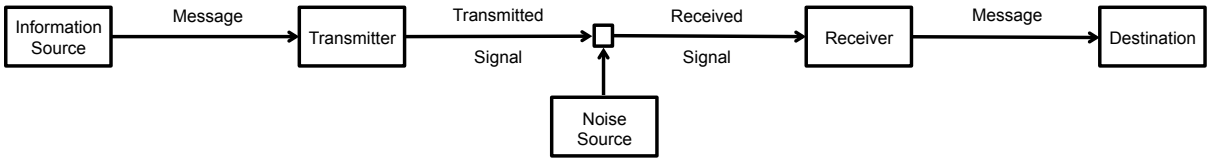


Figure 1: *Classical model of a general communications system* (Shannon, 1948)*. The noise source can have any statistical distribution, such that this model can accommodate and be applied to any form of neural communications channels*.

At the transmitter (transmitting neuron): The information source is assumed have M possible coding (amplitude) levels. Its precision is varied from 2 coding levels (M=2) to 1001 (M=1001, having 3 decimal places of precision) in order to examine the effects of noise as the number of coding levels increases. The case of M=1001 levels serves as our proxy for continuous representation. Here, for simplicity, we assume amplitude modulation, with amplitude levels (coding levels) normalized and equally spaced between 0 and 1 (i.e. [0,1]). Note that, even though we are simulating the effect of numerous coding levels, the approach is extendable to other modulation schemes (Proakis, 2000; Haykin, 2001; Vitetta et al., 2013), such as frequency modulation (normalized 0 is the minimum frequency and normalized 1 the maximum frequency).

The channel and noise source: Rushton (1961) modeled both the spike signal degradation and the boosting due to each myelin-Node of Ravier (NoR) segment as a constant multiplicative noise (i.e. each myelin-NoR segment has the same multiplicative degradation factor). For example, a 1% degradation/boosting error is modeled as 0.99^n, where n is the number of myelin-NoR segments; the signal degrades multiplicatively after each segment, such that if n were of indefinite length, the signal would eventually vanish. However, degradations and boostings along the axon are actually stochastic, something his approach did not take into account. Our work extends and refines Rushton's approach by introducing stochasticity, where each myelin-NoR segment is modeled as an additive Gaussian noise source as in Figure 2. This additive noise approach is appropriate given that the cable model of the axon (Hodgkin & Huxley, 1952) consists of capacitors and resistors (additive circuitry), instead of transistors/amplifiers (multiplicative circuitry). This adaptation allows us to simulate the cumulative effect of stochastic noise during information transmission along the axon. We note here that, even though the input to the channel (the transmitted signal) is discrete (i.e. possessing M possible coding levels, ranging from 2 to 1001), the channel itself is continuous due to the nature of the noise (continuous Gaussian distribution). In this sense, the information is represented in a discrete form, but is communicated over a continuous channel, akin to Morse Code (discrete information) transmitted over a continuous channel (electrical signals, or radio signals).

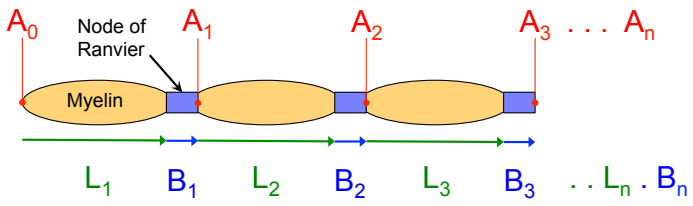


Figure 2: *The additive model of neuronal communications. An axon is modeled as a chain of myelin segments delimited by NoR. As the transmitted signal,* A_0*, propagates along the myelin, it is degraded by a conduction loss,* L_i*, with a Gaussian distribution* $N(-\mu_L,\sigma_L^2)$ *where* $-\mu_L$ *is the average conduction loss and* σ_L^2 *arises from thermal*

noise. At each NoR, the signal is boosted back up by B_i*, with a Gaussian distribution* $N(+\mu_L, \sigma_B^2)$ *where* $+\mu_L$ *is the average boosting level and* σ_B^2 *arises from imperfect (noisy) boosting. After traversing one myelin-NoR segment, the signal,* A_1*, is represented by* $A_1 = A_0 + L_1 + B_1$*, having a Gaussian distribution* $N(A_0 - \mu_L + \mu_L, \sigma_L^2 + \sigma_B^2) = N(A_0, \sigma_L^2 + \sigma_B^2)$*. After traversing n myelin-NoR segments along the axon, the signal becomes* $A_n = A_{n-1} + L_n + B_n$*, with a Gaussian distribution* $N(A_0, n(\sigma_L^2 + \sigma_B^2))$*, which is the signal that arrives at the receiver/destination neuron (i.e. the received signal).*

At the receiver (receiving neuron): Due to noise, the received signal differs from the transmitted signal. Given the received signal, signal detection theory (Peterson et al., 1954) is employed at the receiver to find the optimum (most likely) estimate of the transmitted signal. For example, with 2 coding levels (binary coding), there are 2 possible coded amplitude levels within the range [0,1], namely a 0 message or a 1 message. If a 0 message is transmitted, and a 0.3023 signal received (due to noise perturbation), then, optimal signal detection (Peterson et al., 1954) will decode this signal as a 0 message (i.e. optimal detection threshold is 0.5). If there were 1001 coding levels in the range [0,1], there are 1001 possible amplitude levels that are used for encoding information to be transmitted. We note that 1001 coding levels in the range [0,1] is equivalent to having 3 decimal places of encoding precision. In this case, if a 0 message is transmitted, and a 0.3023 signal received, then, optimal signal detection would decode this as a 0.302 message (since 0.302 is one of the 1001 legitimate encoding levels), thereby resulting in a communications error (a 0 message was transmitted but a 0.302 message decoded) due to noise perturbation. The signal detection theory employed here (Peterson et al., 1954) is the same as that employed in psychology and neuroscience (Tanner & Swets, 1954). It is also the same theory employed in psychophysics to estimate the psychometric functions of Alternative Force Choice (AFC) tasks with M possible choices (i.e. M-AFC tasks, where M>1) (Kingdom & Prins, 2010).

RESULTS

Noise accumulation after n myelin-NoR segments: We first simulated the effect of noise accumulation as a function of the number of myelin-NoR segments, n. In Figure 3, we show examples of how the amplitudes of two transmitted signal levels ($A_0 = 0.2$; $A_0 = 0.8$) may change across multiple myelin-NoR segments. We chose the noise variance levels as $\sigma_L^2 = \sigma_B^2 = 0.0001$ for convenience. In both graphs, we can see that the actual signal level after n myelin-NoR segments (shown as blue dots) deviates significantly from the transmitted level, A_0 (shown as a red dashed line). For the case of $A_0 = 0.2$ (i.e. top graph), the signal is very close to zero after 100 myelin-NoR segments, due to the accumulation of thermal and boosting (or relaying) noise. In the case of $A_0 = 0.8$ (bottom graph), the signal remains close to one after 100 myelin-NoR segments. Note that the graphs/results shown are merely 2 of the many possible outcomes of the transmitted signal due to the stochastic nature of the noise. The key conclusion is that, due to noise accumulation, the amplitude of the received signal can be very different from that of the transmitted signal.

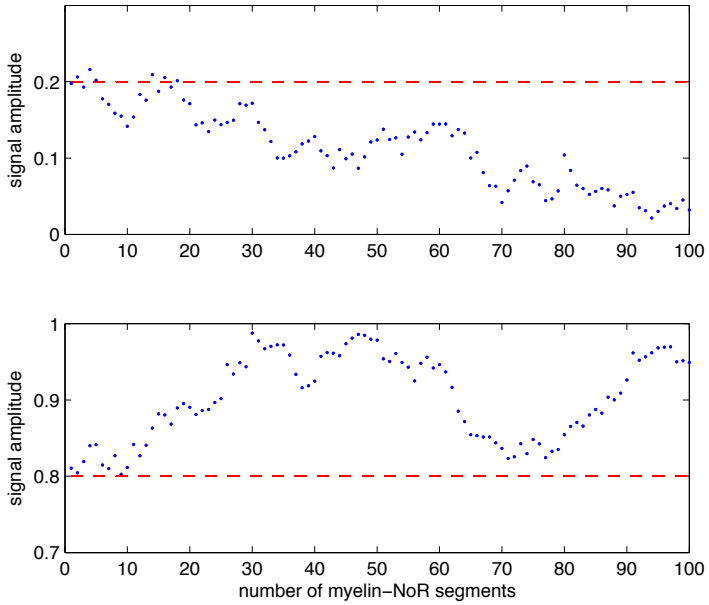


Figure 3: *Two examples of the effect of noise accumulation on the transmitted signal after n myelin-NoR segments. The red dashed lines are the transmitted signals, whereas the blue*

dots are the actual signals after n myelin-NoR segments. Top graph) Results for the case of transmitted signal $A_0 = 0.2$. *Bottom graph) Results for the case of transmitted signal* $A_0 = 0.8$.

Next, we simulated a transmitted signal of level $A_0 = 0.5$ sent repeatedly along the axon (100,000 times). Each column in Figure 4 shows the histogram of the received signals after 1, 10 and 100 myelin-NoR segments respectively. Each row in Figure 4 represents a different level of thermal and/or boosting noise (0.0001, 0.001 and 0.01). Again, for convenience, we assumed that $\sigma_L^2 = \sigma_B^2$. The vertical red lines signify the transmitted signals of level $A_0 = 0.5$. In the first row (lowest noise variance), we see that the width of the histogram grows increasingly wider, such that the histogram after 100 segments has a very large spread compared to the transmitted signal (vertical red line). In the second row (noise variance = 0.001), the histogram looks similarly wide after only 10 segments. After 100 segments, the histogram looks essentially like a uniform distribution, which is very different from the transmitted signal (vertical red line). In the third row (noise variance = 0.01), we see that the histogram has a wide spread after only 1 segment. After 10 segments, the histogram looks like a uniform distribution. We note that a uniform distribution implies that the received signal is a uniform random number between 0 and 1, even though each of the 100,000 transmitted signals has an identical level of 0.5. Thus, due to the accumulation of noise, the distribution of the received signals is very different to that of the transmitted signal. To provide perspective on what 100 myelin-NoR segments translate to in terms of the nervous system, there is an estimated 800 myelin-NoR segments between a person's finger and the spinal cord (Rushton, 1961).

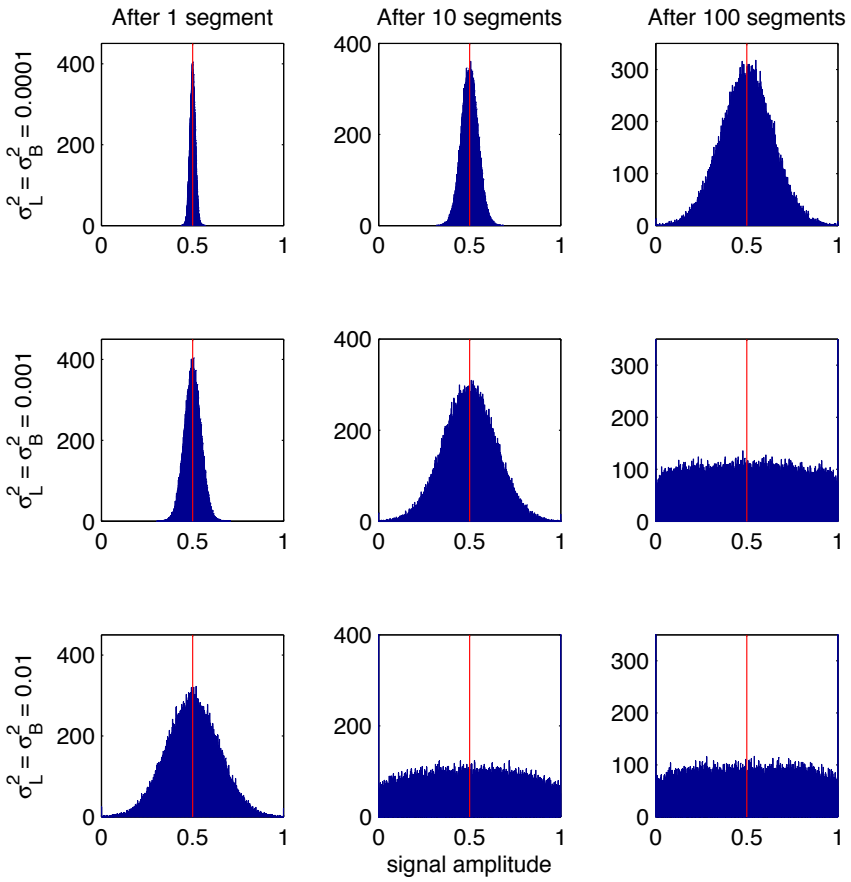


Figure 4: *Histogram of the received signals for the transmitted signal of level* $A_0 = 0.5$ *(i.e. vertical red lines) which was communicated repeatedly 100,000 times along the axon. Each column represents the histogram of the received signals after 1, 10 and 100 myelin-NoR segments respectively. Each row represents a different level of thermal and boosting noise (i.e. 0.0001, 0.001 and 0.01). For convenience, we assume that* $\sigma_L^2 = \sigma_B^2$.

Quantifying information degradation: Searching the literature, we were unsuccessful in locating previous work in quantifying how information degrades along the axon. The contribution of the present work lies in the use of Shannon mutual information (Cover, 2006) to quantify how the input information degrades across myelin-NOR segments along the axon (output). We may write the mutual information of any channel (of input X and output Y) as:

$$I(X;Y) = H(X) - H(X \mid Y)$$

where $H(X)$ is the marginal entropy and $H(X \mid Y)$ is the conditional entropy. For ease of comparison, we use the normalized mutual information (NMI) (Strehl & Ghosh, 2002):

$$NMI = \frac{I(X;Y)}{\sqrt{H(X)H(Y)}}$$

Here, an NMI value of 1 means that no information has been lost, whereas an NMI value of 0 means that all information has been lost.

Figure 5 shows the NMI after n myelin-NoR segments for various numbers of coding levels ranging from 2 to 1001. As above, we utilize the case of 1001 coding levels as a proxy for continuous representation. At one extreme, for the case of 2 coding levels, NMI remains almost unchanged (i.e. 0.9952) after 100 segments: almost no information is lost. As the number of coding levels increase, NMI drops noticeably. At the other extreme, for 1001 coding levels, NMI decreases to 0.0027 after 100 segments: almost none of the original transmitted information remains. In fact, NMI drops from 1 to 0.1148 after only 1 segment, and 0.0203 after 10 segments. It is obvious that, for more than 1001 coding levels (nearly continuous), information degradation increases. The key result is that information is lost rapidly when a continuous representation is employed. For a discrete representation, fewer coding levels lead to lower and slower information loss.

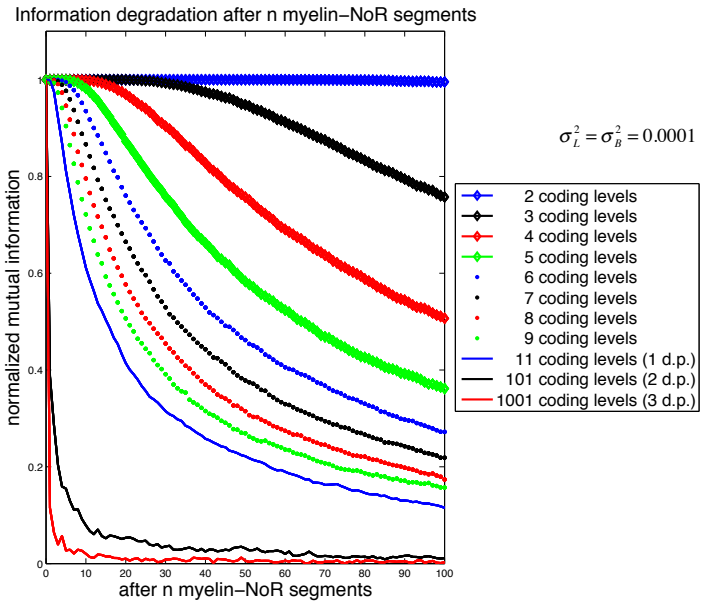


Figure 5: *Information degradation after n myelin-NoR segments for different numbers of coding levels, ranging from 2 to 1001 (i.e. 3 decimal places). For a fair comparison, all noise levels are fixed at 0.0001.*

Noise accumulation after N neurons: We extended our simulation to transmission through multiple successive neurons, where each is modeled as consisting of 100 myelin-NoR segments. Figure 6 shows the information degradation after N neurons. Not surprisingly, for the case of 1001 coding levels (our proxy for continuous representation), information is almost completely lost after 10 neurons (NMI = 0.0002). For the discrete case of 2 coding levels, NMI remains at 0.9676 after 10 successive neurons. The results here agree with those in the previous graph; information is lost rapidly when a continuous representation is employed, and fewer coding levels lead to much less rapid information loss.

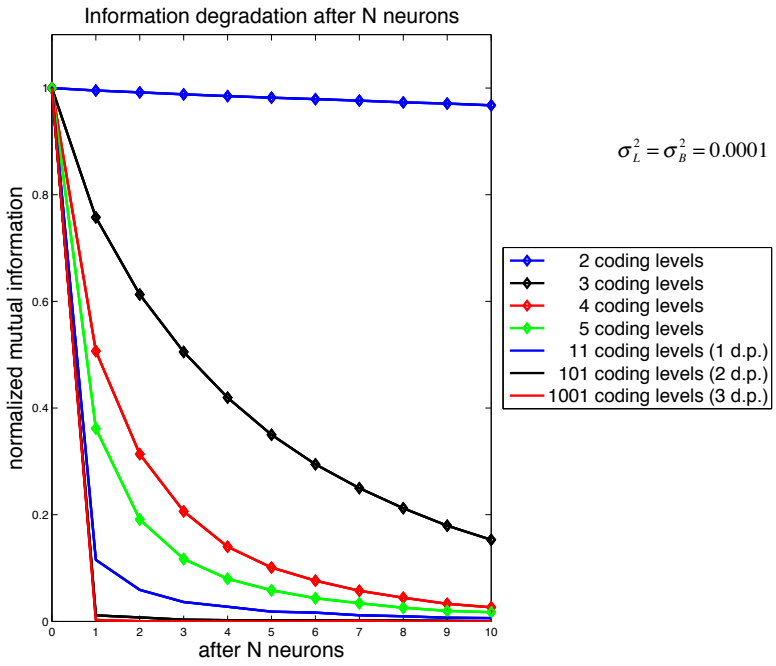


Figure 6: *Information degradation after N neurons.*

INTERIM DISCUSSION

A complete theory of mind/brain must include both computation and communication. Our work demonstrates that, in the presence of noise, it is impossible to communicate reliably between neurons under repeated transmissions using continuous representation. This result provides a neurobiological basis to support the discrete representation of information in the brain. We emphasize that there is currently no consensus in neuroscience on how spikes convey information. Nevertheless, our discrete conclusion here is generalizable to most commonly theorized forms of neural coding schemes, such as spike-count rate coding (Dayan & Abbott, 2001), interspike interval (ISI) coding (Rieke et al., 1997) and Recurrent Neural Network models (McCulloch & Pitts, 1943; Zhang et al., 2014; Yuste, 2015). For example, in the case of spike-count rate coding, suppose that a neuron were to convey a message encoded as 20 spikes/sec. Due to precision limitations and noise, a neuron cannot consistently send exactly the same 20 spikes/sec signal each

time; on one occasion, it may send 19 spikes/sec, while on another occasion, it may send 22 spikes/sec. It is then plausible that all rates between, say, 19 and 22 spikes/sec convey the same message in discrete form, as opposed to the continuous case where 19 spikes/sec conveys a different message to 19.5 spikes/sec. In subsequent sections of this paper, we further explore our discrete hypothesis using behavioral experiments.

Part 2: A Psychophysics (Behavioral) Experiment on Probability Estimation

Humans have a tendency to overestimate small probabilities (dying in an airplane crash, winning the Powerball jackpot lottery) and underestimate large probabilities (e.g. getting lung cancer from smoking cigarettes) when making decisions under risk and uncertainty. These subjective probabilities are typically modeled using a probability distortion (or weighting) function (Tversky & Kahneman, 1992; Camerer & Ho, 1994; Prelec, 1998; Gonzalez & Wu, 1999). It is commonly assumed that subjective probability is represented in the brain in continuous form (a Real number that can take any value between 0 and 1). Such an assumption may result in inaccurate analytical findings if participants' decisions are made using a non-continuous (finite precision) representation. For example, if a participant were to choose between a $100 lottery of probability 0.89721, versus, a $150 lottery of probability 0.50219, the participant will typically round the numbers when making the choice (Kahneman & Tversky, 1979). By doing so, the choice is greatly simplified to a $100 lottery with probability 0.9, versus, a $150 lottery with probability 0.5. Then, analysis performed using a continuous representation is, strictly speaking, inaccurate because 0.89721 is actually treated indifferently to 0.9. Rounding implies that a range of probabilities is treated as being the same (indifferent). This inconvenient problem is often ignored because standard (continuous) models cannot easily

take into account such indifference. While the work of Kahneman and Tversky (1979) expected rounding (i.e. "editing"), they did not quantify it.

Quantization (Gray & Neuhoff, 1998) is the term used to describe an encoding process in which a continuous quantity (a Real number or analog signal) is systematically divided into a finite number of possible categories. We note that the term "quantization" is loosely synonymous to terminologies such as discretization, chunking, categorization and classification. The term "discrete" is common in neuroscience, whereas the term "quantized" is more common in engineering. Here, we use both terms interchangeably. Rounding a number is the oldest example of quantization (Sheppard, 1898). Oliver et al. (1948) employed quantization to convert continuous signals (e.g. audio/voice) to their digital (i.e. binary) forms to enable the maximum possible efficiency of data compression and communications (Shannon, 1948), laying the theoretical foundations for almost all communications systems in use today. For binary quantization, the total number of categories equals to 2^n, where n is the quantization precision expressed in bits. For example, suppose that we are encoding probabilities (Real numbers in the interval [0,1]). If n were one, then we have only two (2^1) output categories. We might encode all quantities in the interval [0, 0.5] as the first output category, with the remaining probabilities (0.5, 1] as the second category. The distinction between substantially different probabilities such as 0.001 and 0.5 is lost in the encoding (i.e. 0.001 and 0.5 are mapped indifferently to the same category). When n is large, such as 24 bits, we can encode to 2^{24} possible very finely spaced categories, resulting in only a very small loss of precision compared to the actual Real number. Thus, a large number of bits represents a higher level of precision (in the engineering sense). A continuous quantity is therefore the special case where $n \rightarrow \infty$. Note that, in typical electrophysiology equipment, continuous neural signals (electrical voltage data in neurons) from electrodes are passed on to a high precision (16-bit) analog-to-digital converter (ADC, also known as a quantizer) before being stored as a digital/discrete data file (Squires, 2013).

Building on Shannon's work (Shannon, 1948), Miller (1956) hypothesized that the magical number 7, plus or minus 2, represents the brain's capacity limit for processing information. He reported experimental evidence showing that the brain's accuracy limit for

absolute judgments of the position of a pointer in a linear interval to be about 3.25 bits, suggesting that the brain categorizes (discretized or quantized) stimuli into 9 or 10 ($2^{3.25}$ = 9.5) discrete categories (instead of a continuous or infinite number of categories). Similar to Miller's work, we hypothesize that the brain has a capacity limit in estimating probabilities. This implies that probability estimates in humans may not be represented as a continuum between 0 and 1, but rather, as a discrete set of categories (say, very low, low, medium, high, and very high), such that certain ranges of distinct probability values (say, from 0.1 to 0.3) will be treated indifferently (discretized into the same "low" category). In this section, we conduct three human behavioral experiments and analyze the experimental data using a novel quantized (i.e. discrete) probability distortion function, in comparison to the conventional/continuous probability distortion function.

A NOVEL QUANTIZED (DISCRETE) MODEL OF PROBABILITY

At present, subjective probability is modeled by several different probability distortion (or weighting) functions (Tversky & Kahneman, 1992; Camerer & Ho, 1994; Prelec, 1998; Gonzalez & Wu, 1999), all of which assume continuous representation. In the present work, we will use the two-parameter continuous function by Prelec (1998), defined as:

$$w(x) = e^{-\delta(-\log x)^{\gamma}}$$

where $\delta > 0$ and $\gamma > 0$. Depending on the values of δ and γ, this function can be either s or inverse-S shaped. Figure 7C shows an example of a continuous Prelec function.

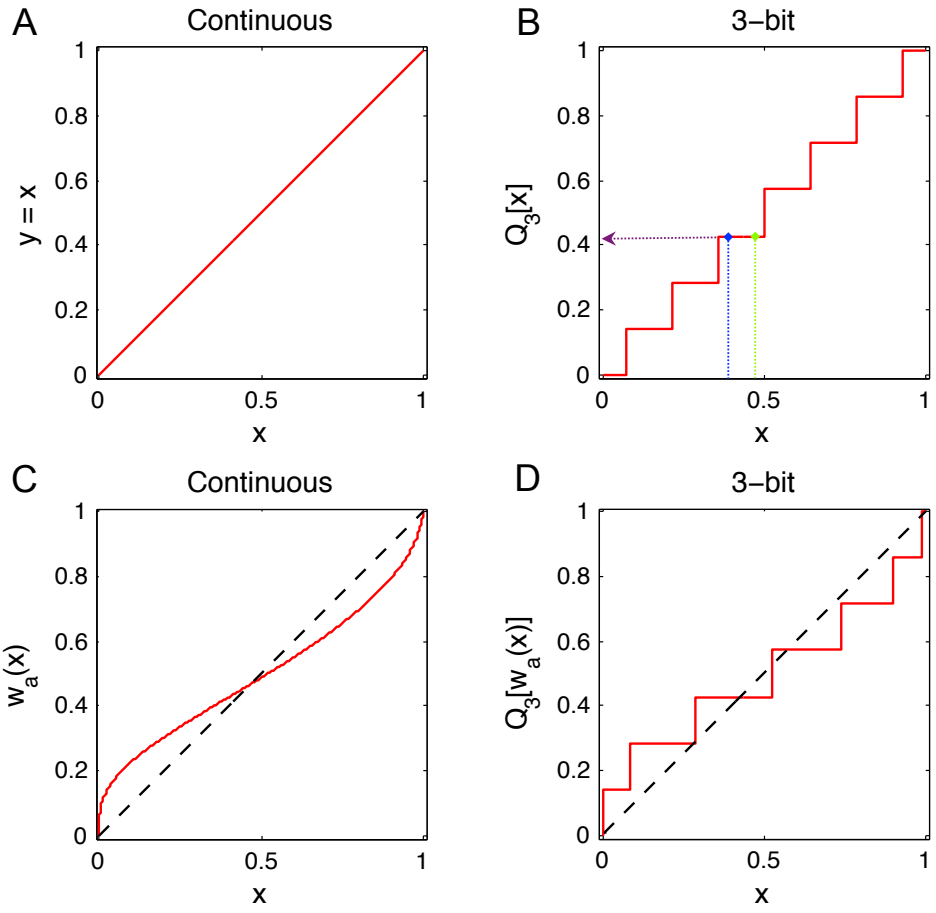


Figure 7: *Examples of continuous functions and their corresponding quantized versions. (A) Continuous linear function y = x. (B) 3-bit quantized version of linear function, $Q_3[x]$. The arrows show the quantization effect, where two different values of x (i.e. 0.39 in blue, and 0.47 in green) both map to the same (indifferent) value of y (i.e. ~0.42 in purple). (C) Continuous Prelec function, $w_a(x)$, with $\delta = 0.9$ and $\gamma = 0.6$. (D) 3-bit quantized version of the Prelec function, $Q_3[w_a(x)]$.*

We quantize (Gray & Neuhoff, 1998) the continuous Prelec function, resulting in the novel quantized (discrete) version:

$$Q_n[w(x)] = Q_n[e^{-\delta(-\log x)^\gamma}]$$

where w(x) is a continuous Prelec function as above, n is the number of bits and $Q_n[.]$ is a quantization function. The quantization function divides the continuous probability space [0,1] into 2^n equally spaced steps, with each step size being equal to 2^{-n} (the step sizes of the corresponding values of n are shown in Appendix 2, Table S1, third column,). Using y = x as a simple example, $y = Q_3[x]$ divides y into $2^3 = 8$ equally spaced steps. These are plotted in Figures 7A and 7B. The quantization (or systematic rounding) effect is further demonstrated in the 3-bit graph (Figure 7B), where it is shown that two different values of x (i.e. 0.39 in blue, and 0.47 in green) both map to the same (indifferent) value of y (~0.42 in purple). In the same manner, a continuous Prelec function, $w_a(x)$, can be quantized to produce $y = Q_n[w_a(x)]$. This quantized function is specified by 3 parameters: δ, γ and n. Figures 7C and 7D show such a continuous Prelec function ($\delta = 0.9$, $\gamma = 0.6$) and its corresponding 3-bit quantized version respectively. We note that n determines the step size, while δ and γ determine where these steps occur. If n is large enough, the quantized Prelec function becomes almost indistinct from the continuous version. A continuous Prelec function is a special case of the quantized Prelec function with $n \rightarrow \infty$.

METHODS

In the following, we consider 3 psychophysics experiments, based on roulette wheels.

Ethical approval: All human experimental procedures were performed at New York University in accordance with the guidelines of the National Institutes of Health and approved by the University Committee on Activities Involving Human Subjects (UCAIHS) serving as New York University's Institutional Review Board (IRB). The experimental protocol was explained to all participants, after which they provided their written informed consent.

Experiment 1: This is a 2-event conjunction task, as shown in Figure 8A. During each trial of the 2-wheel task, a participant has up to 10 seconds to choose between a single roulette wheel with probability r of success and a pair of independent roulette wheels with probabilities x and z of success. The participant's estimates of x, z, and r were based on visual judgments of the fraction of each roulette wheel colored gold. After a participant has made his/her choice, the wheels were spun simultaneously at random speeds, and the participant won a monetary prize if he: (i) chose the single wheel and it stopped in the gold, or, (ii) chose the pair and both stopped in the gold. A staircase procedure (Cornsweet, 1962; Wetherill & Levitt, 1965) was used to estimate the r for which the participant chose the pair as often as the single wheel: r ~ (x,z). Twelve experimental conditions (i.e. x,z pairs) were presented (see Appendix 3, Figure S1A). By design, x is always greater than or equal to z.

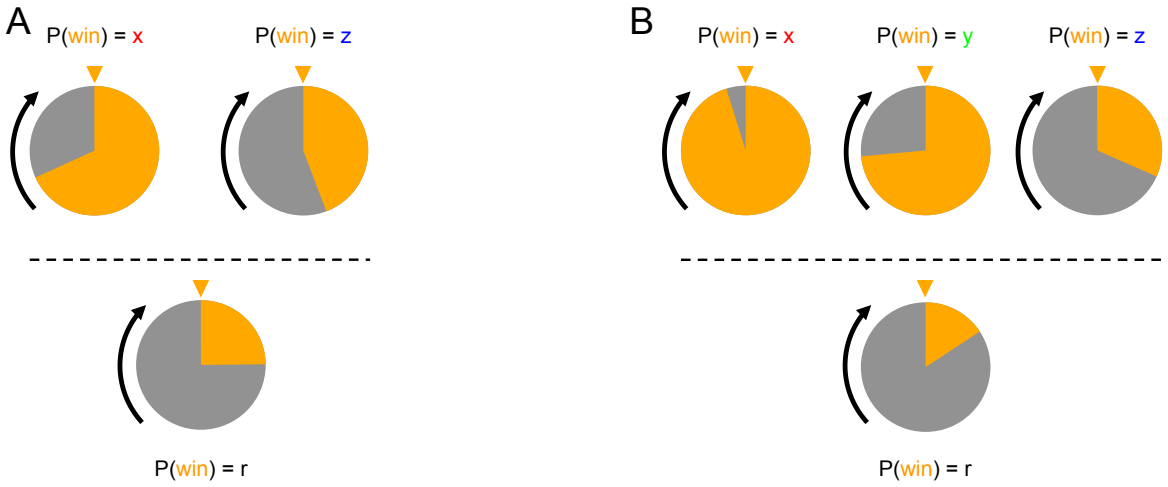


Figure 8: *A) The 2-event (or 2-wheel) conjunction task. B) The 3-event (or 3-wheel) conjunction task of experiment 3.*

Experiment 2: Here, we ran the same design as experiment 1, except for the experimental conditions (x,z pairs) where we employed a different set of 15 experimental conditions (see Appendix 3, Figure S1B).

Experiment 3: The design setup for this experiment is identical to the previous experiments, except that it is a 3-event conjunction task (see Figure 8B). Twenty experimental conditions were employed (see Appendix 3, Figure S1C). By design, $x > y > z$.

Participants: There was a total of 87 distinct participants: 23 participants for experiment 1; 21 for experiment 2; and 44 for experiment 3; one participant participated in both experiments 1 and 3. All participants were naïve, and aged between 18 and 36 years old. There were 53 females and 34 males. Participants were paid $12 per hour. At the end of each experiment, a previously completed trial was randomly drawn, and if this trial was one that the participant had previously won, a $10 bonus was paid.

Measurement and stimuli presentation: We used a one-up/one-down staircase procedure Cornsweet, 1962; Wetherill & Levitt, 1965) with 50 trials per experimental condition such that experiments 1, 2 and 3 consisted of 600, 750 and 1,000 trials respectively. In each experiment, successive conditions were presented in a random order. The horizontal order of the "large" (i.e. x), "midsize" (i.e. y) and "small" (i.e. z) wheels was randomized to mitigate order effects. The vertical placement (top or bottom) of the staircase wheel (r) and the conjunction wheels were also randomized. 30-second mandatory breaks were spaced up uniformly throughout the experiment; experiment 1 had 9 breaks; experiment 2 had 10; experiment 3 had 13. The experiments were implemented using Psychtoolbox (Brainard, 1997; Pelli, 1997), carried out on a 19" touchscreen powered by Windows XP.

Conjunction fallacy: When two events, X and Z, are independent, their conjunction (2-event) probability is the product of the constituent probabilities: $P(X \& Z) = P(X)P(Z)$. Mathematically, a conjunction probability cannot be greater than the probability of its constituents, such that $P(X \& Z) \leq P(X) \leq P(Z)$. In the "Linda the bank teller" study of Tversky and Kahneman (1983), they found that 85% of participants (i.e. 75/88 participants) violated this ordinal rule, thereby committing conjunction fallacy. To investigate conjunction fallacy, we used a one-tailed binomial test against the ordinal rule

at a 0.05 level of significance.

Data fitting: Maximum likelihood estimation (Myung, 2003) was employed to fit the experimental data to each model. All estimations were performed in Matlab at double precision (Matlab's default).

Nested hypothesis tests: Nested hypothesis tests (Mood et al., 1974) were conducted at a 0.05 level of significance. For the 2-wheel task of experiments 1 and 2, we used two quantized Prelec functions to model the conjunction, $r = Q_k[w_a(x)]Q_n[w_c(z)]$, where $Q_k[w_a(x)]$ and $Q_n[w_c(z)]$ are the quantized Prelec functions for the conjunction pair (for wheel x, k is the precision and w_a is the continuous Prelec function; for wheel z, n is the precision and w_c is the Prelec function). This model was applied to a hierarchical sequence of nested hypothesis tests to determine whether:

1. the participants were distorting x and z at all (homogeneous linear case, where estimated probabilities are objective, not subjective);
2. both x and z were subjectively distorted using the same Prelec function (homogeneous Prelec case);
3. x and z were subjective distorted using different Prelec functions (non-homogeneous Prelec case).

All nested models employed for the 2-wheel task of experiments 1 and 2 are summarized in Appendix 4 (Figure S2A). For the 3-wheel task of experiment 3, we used three quantized Prelec functions to model the conjunction, $r = Q_k[w_a(x)]Q_m[w_b(y)]Q_n[w_c(z)]$, where $Q_k[w_a(x)]$, $Q_m[w_b(y)]$ and $Q_n[w_c(z)]$ are the corresponding quantized Prelec functions for the conjunction triplet. A similar hierarchical sequence of nested hypothesis tests was applied. All nested models employed for the 3-wheel task are summarized in Appendix 4 (Figure S2B).

RESULTS

Experiment 1: Three out of 23 participants repeatedly committed conjunction fallacy. For the rest of this paper, all conjunction fallacy participants are excluded. Data fitting is applied to the remaining 20 subjects. Figure 9A plots a heatmap of the negative log likelihood of the maximum likelihood estimation process for one sample participant. Precisions range from 1 to 10 bits, with darker shades showing inferior fits and lighter shades superior fits. On the heatmap, the highest precision is located on the bottom right corner where both x and z are 10 bits. We note that, perceptually, 10 bits of precision has very fine step sizes of $2^{-10} \approx 0.001$ (see Table S1). This corresponds to about 0.35 degrees on the wheel (see Table S1), which is almost indiscernible to the naked eye. As the precision decreases (i.e. moving from the bottom right corner toward the middle), the fit improves. The best fit occurs where x and z are both 5 bits (i.e. white square). This trend offered the first clue that a quantized model may be a better fit than the continuous model.

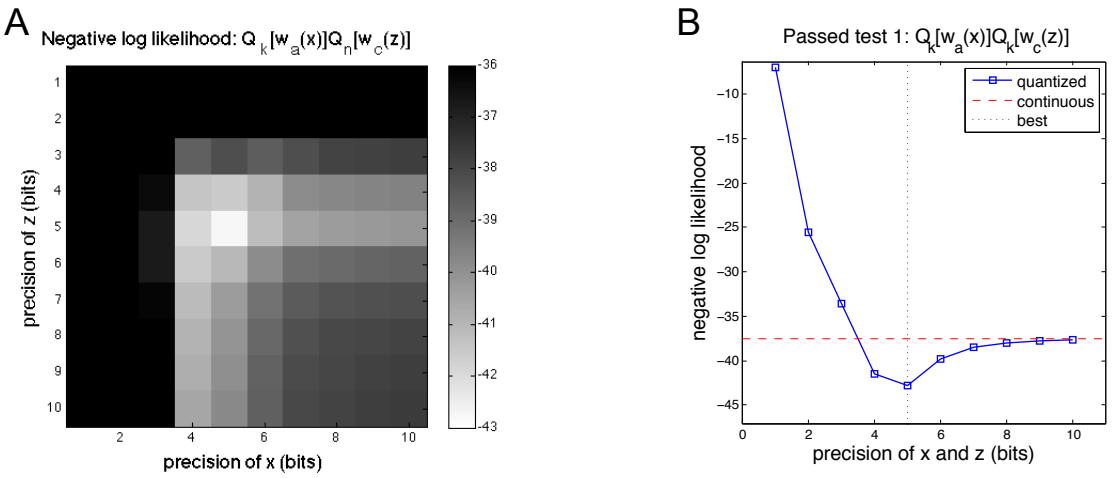


Figure 9: *Negative log likelihood of model fits for one sample participant for the 2-wheel task of experiment 1. (A) The heatmap for the 6-parameter $Q_k[w_a(x)]Q_n[w_c(z)]$ model prior to the nested hypothesis tests. (B) The graph for the 5-parameter $Q_k[w_a(x)]Q_k[w_c(z)]$ model after passing nested hypothesis test 1, showing best fit at 5 bits of precision.*

In terms of nested hypothesis tests, none of the 20 participants passed the linear case (no subjective distortions), 6 participants are homogeneous (same Prelec function for x and z) and 14 participants are non-homogeneous (different Prelec functions for x and z). Details are included in Appendix 5 (Table S2). Interestingly, 17 participants had the same precision for x and z ($Q_k[.]Q_k[.]$ models). The negative log likelihood for the same sample participant is plotted in Figure 9B. This participant passed the test for non-homogeneous Prelec distortion with x and z having the same precision (i.e. $Q_k[w_a(x)]Q_k[w_c(z)]$), which is why the plot is a line instead of a heatmap. The dashed horizontal red line shows the negative log likelihood value for the continuous model. As the precision decreases, the fit improves, with the best fit occurring at 5 bits. This characteristic shape of the negative log likelihood for the quantized case is found for all participants.

The distributions of the best fit precisions of wheels x and z for all participants are plotted in Figures 10A and 10B respectively, confirming that the quantized models fit the experimental data better than the continuous models. The best fit precisions range from 2 to 5 bits, with modes of 4 bits. If deviations from the mode are not significant, a model whereby the precision of the Prelec function is held constant could be a good fit. Such a model would have one less free parameter (allowing one extra degree of freedom). We note that a two-parameter continuous Prelec function is a special case of the three-parameter quantized Prelec function with the precision fixed at ∞. Therefore, by fixing the precision at 4 bits instead of allowing it to be a free parameter (i.e. $Q_4[.]Q_4[.]$ instead of $Q_k[.]Q_k[.]$), we remove one free parameter from the function. We repeated the nested hypothesis test with three additional 4-bit models included in our hypotheses (nested hypothesis test 2) and the results are shown in Appendix 5 (Table S3). Overall, nested hypothesis test 2 produced two key findings: 70% of participants (i.e. 14/20) were non-homogeneous; 70% of participants (i.e. 14/20) were best fit to 4-bit models. The typical probability distortion curves of a homogeneous and non-homogeneous participant are shown in Figures 10C and 10D respectively. For further interpretations of homogenous and non-homogenous participants, see Appendix 6.

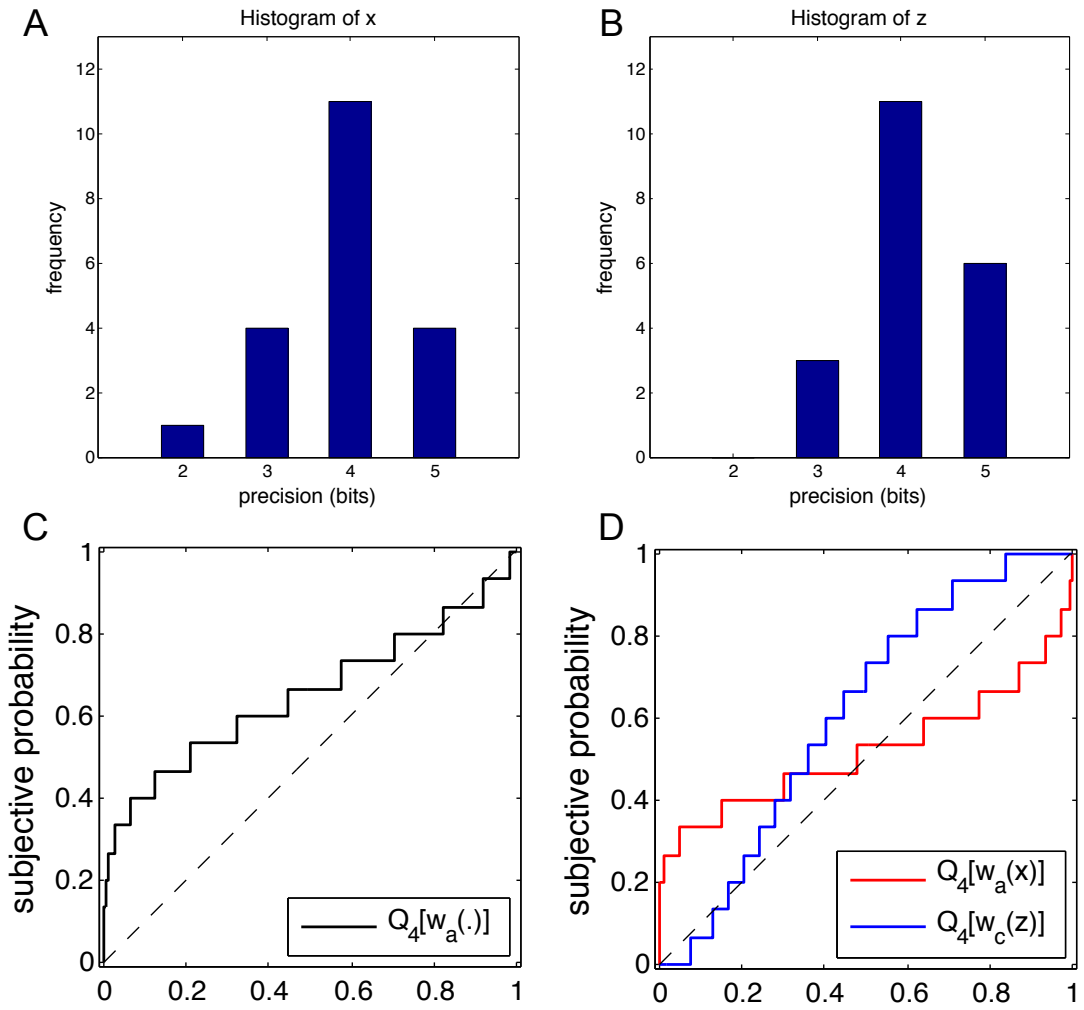


Figure 10: *Distributions of best fit precisions, and typical probability distortion curves. (A, B) Distributions of best fit precisions of wheels x and z for all participants from experiment 1 after passing nested hypothesis test 1. This confirms that quantized models do fit the experimental data better than continuous models, with a mode of 4 bits for both wheels x and z. (C, D) The probability distortion curves of typical homogeneous and non-homogeneous participants of experiment 1.*

Experiment 2: Five out of 21 participants committed conjunction fallacy, which is consistent with experiment 1. The same data fitting and nested hypothesis tests were applied to the remaining 16 participants. Nested hypothesis tests results (detailed in

Appendix 5, Table S4) resemble those of experiment 1: 63% of participants (10/16) were non-homogeneous; 94% of participants (15/16) were best fit to 4-bit models. These results, along with probability distortion curves (not shown), are consistent with those of experiment 1, which tells us that the results of experiment 1 were generalizable and not specific to its experimental conditions.

Experiment 3: Four out of 44 participants committed conjunction fallacy, which is consistent with the findings from experiments 1 and 2. Similar data fitting and nested hypothesis tests were applied to the remaining 40 participants. Nested hypothesis tests results (detailed in Appendix 5, Table S5) resemble those of the 2-wheel experiments: 55% of participants (22/40) were non-homogenous; 75% of participants (30/40) were best fit to 4-bit models. Typical probability distortion curves of a homogeneous and non-homogeneous participant are shown in Appendix 6 (Figures S3B and S3D).

INTERIM DISCUSSION

To summarize, Figure 11 compares the best-fit 4-bit quantized probability model (in black) with the best-fit continuous probability model (in red) for one sample participant. The stimuli (wheels of experiment) are plotted as dots. The gap between the black line and the red curve represents the effect size (the difference between the quantized model versus the continuous model). For some stimuli conditions, the effect size is small (e.g. highlighted in blue), whereas for other stimuli conditions, the effect size is quite large (e.g. highlighted in green). A visual intuition for quantized probability is further illustrated in Appendix 7. Regardless of whether the effect sizes are small or large, the 4-bit quantized model makes a substantially different prediction in human subjective probability when compared to the continuous model.

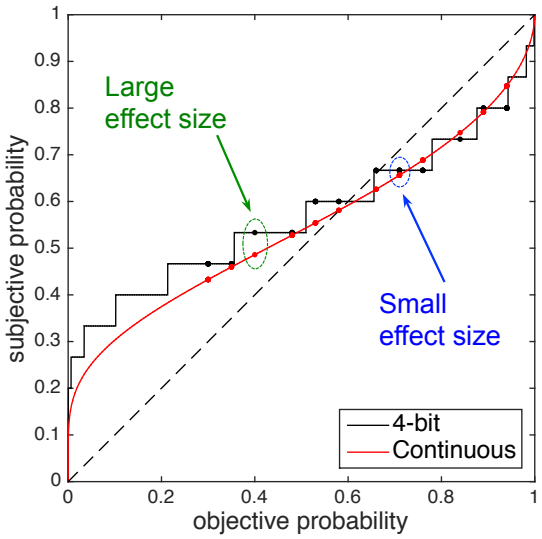


Figure 11: *Illustration of the effect sizes between the best-fit 4-bit model and the best-fit continuous model for one sample participant.*

Overall, 78% of all participants (59/76) were best fit to 4-bit probability models, having 4 bits of quantization per wheel (see Appendix 8 for a summary of all model parameters). On the basis of these results, we conclude that probability is almost certainly represented in a discrete manner in the brain, and very likely to be at 4 bits of precision, as opposed to the commonly assumed continuous representation.

Part 3: A Re-analysis of a Published Behavioral Experiment on Intertemporal Choice

Intertemporal choice, also known as discounting, focuses on value decision trade-offs at different points in time. For example, would you prefer to receive a $10 payment today (present option), or wait for now and receive a $15 payment next month (future option)? Experimental data on intertemporal choices are typically modeled using either a hyperbolic

discounting function (Mazur, 1987; Frederick et al., 2002); Green & Myerson, 2004) or an exponential discounting function (Laibson, 1997; Green & Myerson, 2004; Chabris et al., 2008). The primary difference between the 2 lies in the steepness of the discounting curves; the hyperbolic function decays at a steeper pace than the exponential discounting function, signifying a value decision preference for the present option, as opposed to a future option. Equivalently, preference for the present option signifies a decision maker who will choose the future option only if the payment amount for the future option is significantly larger (say, $20) than the present option (say, $10). To date, discounting functions have been modeled in terms of continuous Real numbers. A quantized (discrete) representation of value was recently proposed (Woodford, 2014). We hypothesize that intertemporal choices (value with a time dimension) are also quantized. For example, people may treat (or discount) 16 days indifferently to 17 days. In this section, we re-analyze the experimental data from Cox and Kable (2014) using novel quantized (discrete) discounting models, and compare them with conventional, continuous discounting models.

A NOVEL QUANTIZED (DISCRETE) HYPERBOLIC DISCOUNTING MODEL

Intertemporal choices are typically modeled using a continuous hyperbolic discounting function (Mazur, 1987):

$$SV = \frac{A}{1+kD}$$
$$\frac{SV}{A} = \frac{1}{1+kD}$$

where A is the objective value, D is the time delay (in unit of days), k is the discount rate and SV is the subjective value. Figure 12 (left) shows an example of a conventional, continuous hyperbolic discounting function.

We quantize (Gray & Neuhoff, 1998) the continuous hyperbolic function, resulting in the form:

$$Q_n\left[\frac{SV}{A}\right] = Q_n\left[\frac{1}{1+kD}\right]$$

where n is the number of bits and Q_n denotes a quantization function that divides the hyperbolic discounting function into 2^n possible steps or quantization levels. Figure 12 (right) shows an example of a 3-bit quantized hyperbolic discounting model (3 bits = 2^3 = 8 levels). We note here that the conventional, continuous model is simply a quantized model with an infinite number of steps (quantization levels).

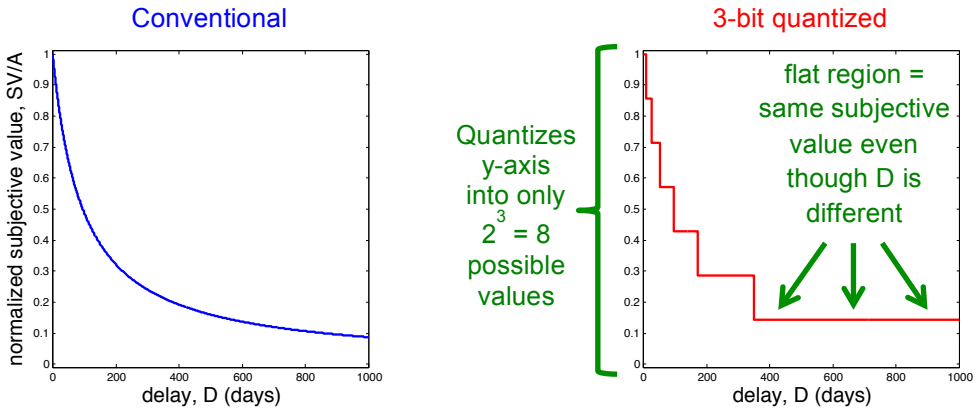


Figure 12: *Left) Conventional, continuous hyperbolic discounting model. Right) 3-bit quantized hyperbolic discounting model.*

The same approach can also be applied to a conventional exponential discounting function (Laibson, 1997; Green & Myerson, 2004; Chabris et al., 2008), to produce a quantized exponential model (see Appendix 9).

METHODS

Our analysis is a re-examination of the human behavioral data previously collected for an fMRI study by Cox and Kable (2014). We first briefly outline their methods, which were approved by the Institutional Review Board of the University of Pennsylvania (Cox

& Kable, 2014). During each trial, a participant chooses between 2 options: $40 now, or $X in D days. $X is capped at $100, while D ranges from 1 to 327 days. There were 204 trials in total. Participants were paid $20 for their participation in the study. At the end of the experiment, one of the completed 204 trials was randomly selected and a bonus corresponding to the participant's choice in the selected trial was paid. For example, if the randomly selected trial was a choice between receiving $40 now (present option) versus receiving $60 in 18 days' time (future option) and the participant had (during the experiment) chosen the present option, then, a $40 bonus was paid to the participant. If the participant had chosen the future option instead, then, a $60 bonus was paid to the participant after an 18-day delay. The bonus was paid using a debit card with the corresponding delay date. A total of 20 participants performed the task. In terms of data analysis, we extended the same maximum likelihood estimation approach for data fitting (as in Cox and Kable, 2014)) using the quantized hyperbolic discounting model. We also employed nested hypothesis testing (Mood et al., 1974), similar to the approach in Part 2 of this paper.

RESULTS

Figure 13 shows the negative log likelihood of the maximum likelihood estimation process for one sample participant. Precision range from 1 to 16 bits. The fit for the continuous model is shown in the horizontal dashed red line. As the quantized precision (i.e. blue line) increases from 1 to 5 bits, the fit improves (value of negative log likelihood decreases). Beyond that, the fit becomes worse (i.e. value of negative log likelihood increases) and subsequently flattens off at the level of the continuous model. The best fit occurs at a precision of 5 bits, suggesting that a quantized model may be a better fit than a continuous model.

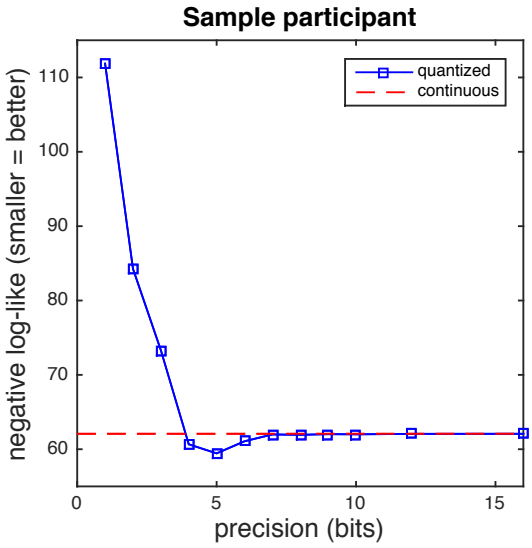


Figure 13: *Negative log likelihood of hyperbolic model fit for one sample participant.*

Of the 20 participants, 9 were best fit to 5-bit quantized hyperbolic models (i.e. 2^5 = 32 steps). The histograms of fitted model parameters are shown in Figure 14. We note that the largest number of bits resulting from the data fitting exercise is 9 bits, representing a model with 2^9 = 512 levels; a continuous hyperbolic discounting model is the case of an infinite number of levels.

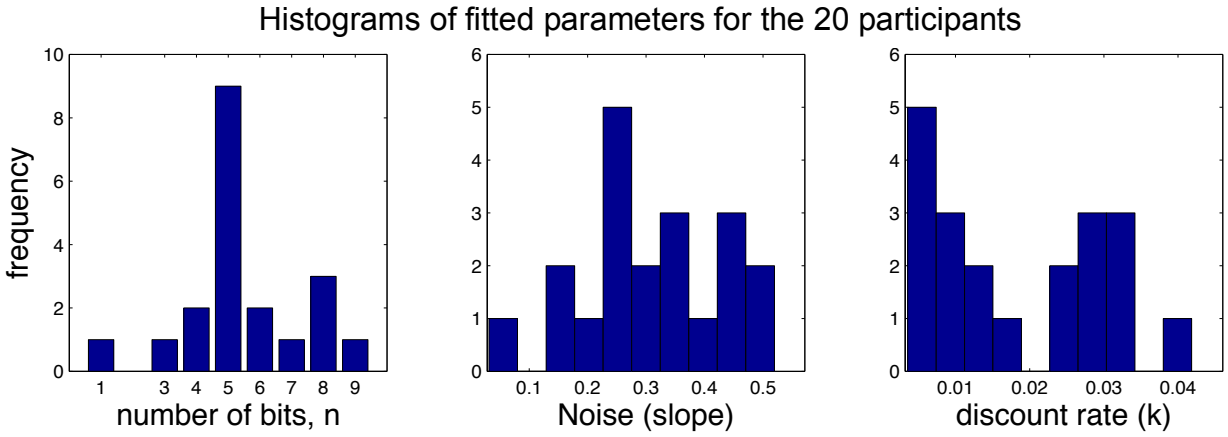


Figure 14: *Histogram of fitted parameters (20 participants) for the quantized hyperbolic model.*

Our quantized hyperbolic discounting model has 2 free parameters (n and k). Since the experimental data has a mode of 5 bits, we applied a nested hypothesis test (Mood et al., 1974) on the model with precision fixed at 5 bits instead of being a free parameter:

$$Q_n\left[\frac{1}{1+kD}\right] \quad \rightarrow \quad Q_5\left[\frac{1}{1+kD}\right]$$

The model on the left has 2 free parameters (i.e. n, k) whereas the model on the right has only 1 free parameter (i.e. k). The purpose of the nested hypothesis test (Mood et al., 1974) is to explore whether the second parameter is statistically justifiable or required for the data fitting of each participant. We note that such a 1-parameter (k only) model is analogous to the conventional, continuous hyperbolic discounting model except that n is being fixed at 5 bits instead of being fixed at infinity. Results of the nested hypothesis test showed that 17 out of 20 participants were best fit to this 1-parameter model (k is a free parameter while n fixed at 5 bits). The 5-bit quantized hyperbolic discounting curves for two representative participants are shown in Figure 15.

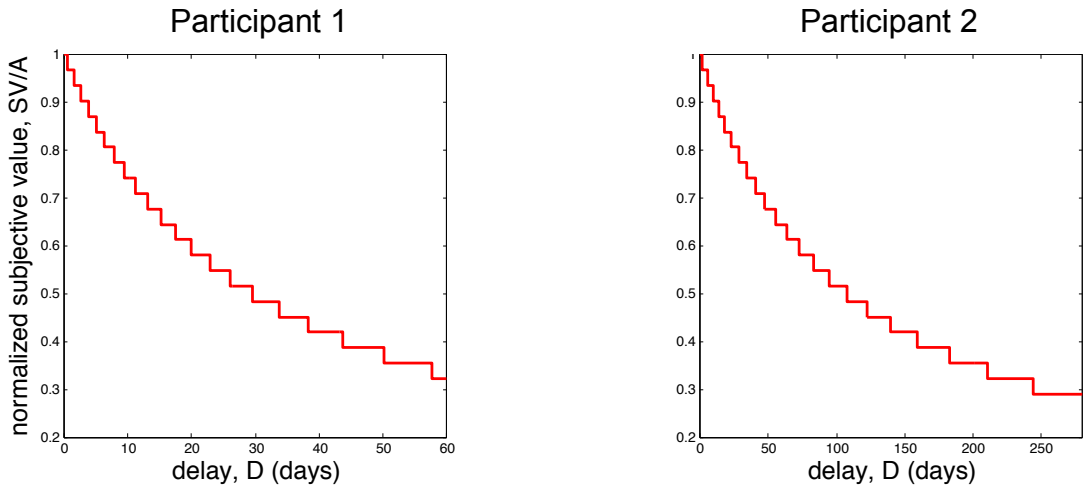


Figure 15: *5-bit quantized hyperbolic discounting curves for two representative participants.*

We also fit the experimental data to a quantized exponential model (see Appendix 10). To summarize, 8 out of the 20 participants were best fit to the 5-bit quantized exponential models. After applying the nested hypothesis test, 15 out of the 20 participants were best fit to the 1-parameter quantized exponential model. These exponential model findings are consistent with the hyperbolic ones.

INTERIM DISCUSSION

For completeness, we compared the performance of the quantized hyperbolic model with the quantized exponential model (see Appendix 11). In summary, 13 out of 20 participants were best fit to a quantized exponential model while the remaining 7 participants were best fit to a quantized hyperbolic model. Overall, this best-of-the-best model comparison concluded that 15 out of 20 participants (75%) have 5-bit precision. A point to emphasize here is that all the quantized models are a better fit to the experimental data than their respective continuous versions. Our re-analysis of the intertemporal choice study by Cox and Kable (2014) confirms our hypothesis that intertemporal choices are indeed quantized. While continuous models have, up till now, been convenient for analyzing experimental data, we must be open to quantized approaches. Given that our quantized result here was attained based on an independent study (a study that was neither designed nor conducted by us), we are confident that our approach is generalizable to many existing and future studies.

Overall Conclusions

In Part 1, our computer simulations showed that information cannot be communicated reliably between neurons using a continuous representation due to the presence of noise. Information representation at the neural level has to be discrete. In Part

2, our extensive behavioral experiment on probability estimation found that data from all participants were better fit to quantized probability models than continuous ones, with 78% of all participants having 4-bit (16 steps) precision. Probability is very likely to be represented in a discrete rather than continuous form in the brain. In Part 3, our re-analysis of an independent study on intertemporal choice by Cox and Kable (2014) reaffirmed the findings of Part 2; that, data from all participants were better fit to quantized models than continuous ones (whether hyperbolic or exponential), with 75% of all participants having 5-bit (32 steps) precision. The independent nature of Part 3 shows the generalizability of our approach to existing and future studies. Overall, the results from Parts 1, 2 and 3 corroborate each other, supporting our discrete hypothesis of information representation in the brain.

We hope that our extensive discrete results here, in conjunction with the macaque LIP discrete results of Latimer et al. (2015), resolve the question of information representation in the brain question, once and for all. Going forward, we firmly believe that the correct research question to explore is no longer that of continuous-versus-discrete, but rather, how fine grained the discreteness is (how many bits of precision). It is very plausible that different parts of the brain (e.g. visual, auditory, cognitive decision-making) operate at different levels of discreteness based on different numbers of quantization levels. For example, camera photographs are commonly encoded in 24-bit RGB color (i.e. 16,777,216 distinct colors) whereas CD music is commonly encoded in 16-bit audio (i.e. 65,536 distinct levels of loudness), representing (or approaching) the maximum limits of the brain's visual and auditory precisions/discreteness respectively.

References

Abbott LF, DePasquale B & Memmesheimer R-M (2016). Building functional networks of spiking model neurons. *Nat Neurosci* 19, 350-355.

Adrian ED (1914). The all-or-none principle in nerve. *J Physiol* 47, 460-474.

Basu K (1997). Why are so many goods priced to end in nine? And why this practice hurts producers. *Econ Lett* 54, 41-44.

Brainard DH (1997). The psychophysics toolbox. *Spat Vis* 10, 433-436.

Busemeyer JR & Diederich A (2014). Estimation and testing of computational psychological models. In *Neuroeconomics: Decision Making and the Brain*, ed. Glimcher PW & Fehr E, pp. 49-61. Academic Press, Cambridge, MA.

Camerer CF & Ho T-H (1994). Violations of the betweenness axiom and nonlinearity in probabilities. *J Risk Uncertain* 8, 167-196.

Chabris CF, Laibson DI & Schuldt JP (2008). Intertemporal choice. In *The New Palgrave Dictionary of Economics Online, Second Edition*, ed. Durlauf SN & Blume LE. Palgrave Macmillan, New York, NY. Available at http://www.dictionaryofeconomics.com/article?id=pde2008_P000365 [Accessed April 29, 2018].

Chaudhuri R & Fiete I (2016). Computational principles of memory. *Nat Neurosci* 19, 394-403.

Cornsweet TN (1962). The Staircase-Method in psychophysics. *Am J Psychol* 75, 485-491.

Cover TM & Thomas JA (2006). *Elements of Information Theory. 2nd edn*. Wiley, Hoboken, NJ.

Cox KM & Kable JW (2014). BOLD subjective value signals exhibit robust range adaptation. *J Neurosci* 34, 16533-16543.

Dayan P & Abbott LF (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press, Cambridge, MA.

Eliasmith CA (2005). A unified approach to building and controlling spiking attractor networks. *Neural Comput* 17, 1276-1314.

Faisal AA, Selen LPJ & Wolpert DM (2008). Noise in the nervous system. *Nat Rev Neurosci* 9, 292-303.

Frederick S, Loewenstein G & O'Donoghue T (2002). Time discounting and time preference: a critical review. *J Econ Lit* 40, 351-401.

Gallistel CR (2016). Presidential column: What we have and haven't learned. *Observer* 29. Available at https://www.psychologicalscience.org/observer/what-we-have-and-havent-learned#.WSFTuhPyuRs [Accessed April 29, 2018].

Gold JI & Shadlen MN (2007). The neural basis of decision making. *Annu Rev Neurosci* 30, 535-574.

Gonzalez R & Wu G (1999). On the shape of the probability weighting function. *Cogn Psychol* 38, 129-166.

Gray RM & Neuhoff DL (1998). Quantization. *IEEE Trans Inf Theory* 44, 2325-2382.

Green L & Myerson J (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychol Bull* 130, 769–792.

Haykin S (2001). *Communications Systems. 4th edn*. Wiley, Hoboken, NJ.

Hodgkin AL & Huxley AF (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol* 117, 500-544.

Hubara I, Courbariaux M, Soudry D, El-Yaniv R & Bengio Y (2016). Quantized neural networks: Training neural networks with low precision weights and activations. arXiv:1609.07061. Available at https://arxiv.org/abs/1609.07061 [Accessed April 29, 2018].

Kahneman D & Tversky A (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263-292.

Khaw MW, Stevens L & Woodford M (2017). Discrete adjustment to a changing environment: Experimental evidence. *J Monet Econ* 91, 88-103.

Kingdom FAA & Prins N (2010). *Psychophysics: A Practical Introduction*. Elsevier, London, UK.

Laibson DI (1997). Golden eggs and hyperbolic discounting. *Q J Econ* 112, 443-477.

Lashley KS (1960). Cerebral organization and behavior. In *The Neuropsychology of Lashley; Selected papers of K.S. Lashley*, ed. Beach FA, Hebb DO, Morgan CT, & Nissen HV, pp. 529-543. McGraw-Hill, New York, NY.

Latimer KW, Yates JL, Meister MLR, Huk AC & Pillow JW (2015). Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science* 349, 184-187.

Lin Z, Courbariaux M, Memisevic R & Bengio Y (2016). Neural networks with few multiplications. *International Conference on Learning Representations*. Available at https://arxiv.org/abs/1510.03009 [Accessed April 29, 2018].

Luck SJ & Vogel EK (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends Cogn Sci* 17, 391-400.

Ma WJ, Husain M & Bays PM (2014). Changing concepts of working memory. *Nat Neurosci* 17, 347-356.

Marr D (1982). *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. W.H. Freeman & Company, San Francisco, CA.

Mazur JE (1987). An adjustment procedure for studying delayed reinforcement. In *The Effect of Delayed and Intervening Events on Reinforcement Value*, ed. Commons ML, Mazur JE, Nevin JA & Rachlin H, pp. 55-73. Erlbaum, Hillsdale, NJ.

McCulloch WS & Pitts WH (1943). A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 5, 115-133.

Miller GA (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol Rev* 63, 81-97.

Mood AM, Graybill FA & Boes DC (1974). *Introduction to the Theory of Statistics*. pp. 440-442. McGraw-Hill, New York, NY.

Myung IJ (2003). Tutorial on maximum likelihood estimation. *J Math Psychol* 47, 90-100.

Nolte J (2008). *The Human Brain: An Introduction to its Functional Anatomy. 6th edn.* Mosby, Maryland Heights, MO.

Oliver BM, Pierce JR & Shannon CE (1948). The philosophy of PCM. *Proc IRE*, 1324-1331.

Pelli DG (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10, 437-442.

Perepelitsa DV (2006). Johnson noise and shot noise. Department of Physics, MIT. Available at http://web.mit.edu/dvp/Public/noise-paper.pdf [Accessed April 29, 2018].

Peterson WW, Birdsall TG & Fox WC (1954). The theory of signal detectability. *Trans IRE Professional Group on Information Theory* 4,171-212.

Prelec D (1998). The probability weighting function. *Econometrica* 66, 497–527.

Proakis JG (2000). *Digital Communications. 4th edn*. McGraw-Hill, New York, NY.

Rescorla M (2016). The computational theory of mind. In *The Stanford Encyclopedia of Philosophy (Winter 2016 edition)*, ed. Zalta EN, Metaphysics Research Lab, Stanford University. Available at https://plato.stanford.edu/archives/win2016/entries/computational-mind/ [Accessed April 29, 2018].

Rieke F, Warland D, van Steveninck RdR & Bialek W (1997). *Spikes: Exploring the Neural Code*. MIT Press, Cambridge, MA.

Rushton WAH (1961). Peripheral coding in the nervous system. In *Sensory Communications*, ed. Rosenblith WA, pp. 169-181. MIT Press, Cambridge, MA.

Shannon CE (1948). A mathematical theory of communication. *Bell Syst Tech J* 27, 379-423, 623-656.

Sheppard WF (1898) On the calculation of the most probable values of frequency constants for data arranged according to equidistant divisions of a scale. *Proc Lond Math Soc* 29, 353–380.

Squires D (2013). Instrumentation electronics for an integrated electrophysiology data acquisition and stimulation system. *Master's Theses* 447, Western Michigan University. Available at http://scholarworks.wmich.edu/masters_theses/447 [Accessed April 29, 2018].

Strehl A & Ghosh J (2002). Cluster ensembles – a knowledge reuse framework for combining multiple partitions. *J Mach Learn Res* 3, 583-617.

Sun JZ, Wang GI, Goyal VK & Varshney LR (2012). A franemwork for Bayesian optimality of psychophysical laws. *J Math Psychol* 56, 495-501.

Tversky A & Kahneman D (1983). Extensional vs. intuitive reasoning: the conjunction fallacy in probability judgment. *Psychol Rev* 90, 293-3l5.

Tanner WP & Swets JA (1954). A decision-making theory of visual detection. *Psychol Rev* 61, 401-409.

Tversky A & Kahneman D (1992). Advances in prospect theory: cumulative representation of uncertainty. *J Risk Uncertain* 5, 297-323.

VanRullen R & Koch C (2003). Is perception discrete or continuous? *Trends Cogn Sci* 17, 207-213.

Varshney LR, Sjöström PJ & Chklovskii DB (2006). Optimal information storage in noisy synapses under resource constraints. *Neuron* 52, 409-423.

Varshney LR & Varshney KR (2017). Decision making with quantized priors leads to discrimination. *Proc IEEE Inst Electr Electron Eng* 105, 241-255.

Vitetta G, Taylor DP, Colavolpe G, Pancaldi F & Martin PA (2013). *Wireless Communications: Algorithmic Techniques*. Wiley, Hoboken, NJ.

Wang XJ (2009). Attractor network models. In *Encyclopedia of Neuroscience vol. 1*, ed. Squire LR, pp. 667-679. Academic Press, Oxford.

Wetherill GB & Levitt H (1965). Sequential estimation of points on a psychometric function. *Br J Math Stat Psychol* 18, 1-10.

Woodford M (2014). Psychophysical aspects of choice behavior. *The Kavli Foundation Social & Decision Science Workshop*, Miami. Available at http://www.neuroeconomics.org/wp-content/uploads/2014/09/Woodford-SNE.pdf [Accessed April 29, 2018].

Yuste R (2015). From the neuron doctrine to neural networks. *Nat Neurosci* 16, 487-497.

Zhang H, Wang Z & Liu D (2014). A comprehensive review of stability analysis of continuous-time recurrent neural networks. *IEEE Trans Neural Netw Learn Syst* 25, 1229-1262.

Additional Information

Acknowledgements: J.T. thanks C. Randy Gallistel and Michael Woodford for their guidance, advice and feedback, which have been instrumental to the success of the work in this paper. J.T. acknowledges the contributions of Laurence T. Maloney on the acquisition of research funding for this work, his inputs on the initial design of the computer

simulation in Part 1 and his inputs on the experimental designs in Part 2. J.T. further acknowledges the contributions of Hang Zhang on her inputs on the experimental designs in Part 2, along with part of its data collection. J.T. thanks Laurence T. Maloney, Denis Pelli, Jonathan Winawer, Joseph Kable and Hang Zhang for their comments and suggestions on this work.

Competing interests: The authors declare no conflict of interest.

Author contributions: The work in this paper is based on J.T.'s PhD dissertation research, undertaken and successfully completed at the Department of Psychology of New York University. J.T. conceived the projects, designed the experiments, carried out the experiments, developed the novel quantized models, analyzed the data, interpreted the results, discussions and conclusions, and wrote this manuscript. D.P.T. contributed to the analysis of the experimental data, the interpretations of the results, discussions and conclusions, and co-wrote the manuscript. All (both) authors approved the final version of this manuscript, and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All (both) persons designated as authors qualify for authorship, and all those who qualify for authorship are listed.

Funding: This work was supported by grant EY019889 from the National Institutes of Health.

Appendix 1: Three Key Aspects of Communications in the Brain

In order to tackle the question of continuous versus discrete representation, we have examined the brain's internal information communications requirements, and now summarize them as the following 3 key aspects of communications:

1) **Information in the brain must be transmitted reliably:** Some information in the sensory system must be transmitted over a relatively long distance while retaining its fidelity. One such example is vision. Information in human vision originates at the retina, is conveyed via the optic nerve to the Lateral Geniculate Nucleus (LGN) and arrives at the visual cortex, located at the back of the brain – a distance spanning almost the entire length of the brain (Nolte, 2008). The fidelity of the visual information must be retained throughout this transmission process. Otherwise, vision will fail.

2) **Information in the brain must be stored and retrieved repeatedly:** One example of this is memory. Our mind needs to be able to remember/recall/retrieve the same information (e.g. date of birth) repeatedly on many occasions possibly throughout an entire lifetime, upon which we would decide on the desired course of action (e.g. how should I celebrate my birthday this year?). Without memory, there is no means to relate to significant information and experiences in our lives. Being able to store and retrieve information repeatedly is crucial to the brain's function.

3) **Noise must exist in the brain:** All electrical processes, including neural spikes, are noisy due to thermal fluctuations, known as thermal noise or Johnson noise (Perepelitsa, 2006). There are additional sources of noise in the brain, such as sensory noise and motor noise (Faisal et al., 2008). Noise affects both information transmission and repeated retrieval of stored information (accessing memory). In the first, noise can degrade/corrupt the information during transmission. In the second, stored information

(memory) could be further compromised by the very act of retrieving/accessing it. That is, accessing memory content risks altering it.

A major consequence of these aspects is that information representation in the brain almost certainly cannot be continuous. If representation were continuous (Real numbers, where each value represents a unique message), then noise would easily cause confusion of one message with another (message1 + noise = message2), no matter how large or small its magnitude (variance).

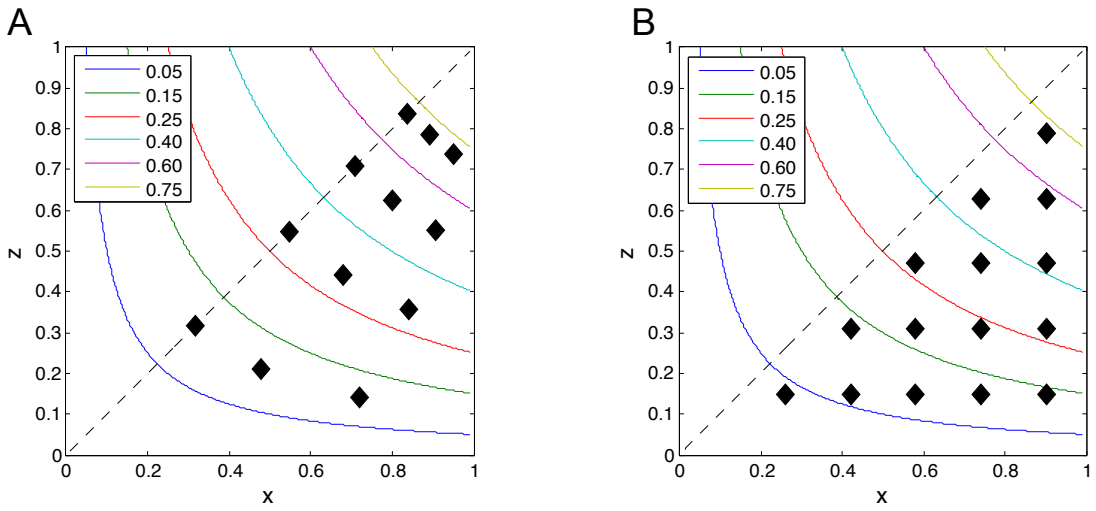
We note that, from a computer engineering standpoint, true Real number computers (i.e. employing continuous representation) are not physically realizable, because there is an uncountable infinity of Real numbers. For example, the memory content of a modern digital computer can be concatenated in order to form a long but finite-length word; whereas true Real numbers have infinite precision (decimal places), and consequently, require infinite space to store even a single Real number (or variable) which violates the basic foundation of information storage. Therefore, from a computer engineering standpoint, a truly continuous computational device is not physically realizable.

Appendix 2: Quantization Step Sizes

Precision (bit)	Number of steps (categories)	Size of each step (category)	Angle on a wheel (degrees)
1	$2^1 = 2$	0.5	180
2	$2^2 = 4$	0.25	90
3	$2^3 = 8$	0.125	45
4	$2^4 = 16$	0.0625	22.5
5	$2^5 = 32$	0.03125	11.25
6	$2^6 = 64$	0.015625	5.625
7	$2^7 = 128$	0.0078125	2.8125
8	$2^8 = 256$	0.00390625	1.40625
9	$2^9 = 512$	0.001953125	0.703125
10	$2^{10} = 1024$	0.0009765625	0.3515625
20	$2^{20} = 1048576$	9.53674×10^{-7}	0.000343323

Table S1: *The size of each step corresponding to the quantization precision.*

Appendix 3: Experimental Conditions



C

x	y	z	Conjunction probability
0.94	0.89	0.84	0.7027
0.94	0.89	0.66	0.5522
0.94	0.89	0.48	0.4016
0.94	0.89	0.30	0.251
0.94	0.71	0.66	0.4405
0.94	0.71	0.48	0.3204
0.94	0.71	0.30	0.2002
0.94	0.53	0.48	0.2391
0.94	0.53	0.30	0.1495
0.94	0.35	0.30	0.0987
0.76	0.71	0.66	0.3561
0.76	0.71	0.48	0.259
0.76	0.71	0.30	0.1619
0.76	0.53	0.48	0.1933
0.76	0.53	0.30	0.1208
0.76	0.35	0.30	0.0798
0.58	0.53	0.48	0.1476
0.58	0.53	0.30	0.0922
0.58	0.35	0.30	0.0609
0.40	0.35	0.30	0.042

Figure S1: *Experimental conditions. (A) The 12 conditions of experiment 1 (i.e. 2-wheel), with $x \geq z$. (B) The 15 conditions of experiment 2 (i.e. 2-wheel), with $x > z$. (C) The 20 conditions of experiment 3 (i.e. 3-wheel), with $x > y > z$.*

Appendix 4: Nested Hypothesis Tests

A

Precision	Distortion function: Homogeneous (i.e. same for x and z)		Distortion function: Non-homogeneous (i.e. different for x and z)
	Linear	Prelec	
Same for x and z	$Q_k[x]Q_k[z]$	$Q_k[w_a(x)]Q_k[w_a(z)]$	$Q_k[w_a(x)]Q_k[w_c(z)]$
Different for x and z	$Q_k[x]Q_n[z]$	$Q_k[w_a(x)]Q_n[w_a(z)]$	$Q_k[w_a(x)]Q_n[w_c(z)]$

B

Precision	Distortion function: Homogeneous (i.e. same for x, y and z)		Distortion function: Non-homogeneous (i.e. different for x, y and z)
	Linear	Prelec	
Same for x, y and z	$Q_k[x]Q_k[y]Q_k[z]$	$Q_k[w_a(x)]Q_k[w_a(y)]Q_k[w_a(z)]$	$Q_k[w_a(x)]Q_k[w_b(y)]Q_k[w_c(z)]$
Different for x, y and z	$Q_k[x]Q_m[y]Q_n[z]$	$Q_k[w_a(x)]Q_m[w_a(y)]Q_n[w_a(z)]$	$Q_k[w_a(x)]Q_m[w_b(y)]Q_n[w_c(z)]$
4 bits for each of x, y and z	$Q_4[x]Q_4[y]Q_4[z]$	$Q_4[w_a(x)]Q_4[w_a(y)]Q_4[w_a(z)]$	$Q_4[w_a(x)]Q_4[w_b(y)]Q_4[w_c(z)]$

Figure S2: *Nested hypothesis tests. (A) Models used in the 2-wheel task of experiments 1 and 2. (B) Models used in the 3-wheel task of experiment 3.*

Appendix 5: Results of Nested Hypothesis Tests

Models	Homogeneous? (i.e. same distortion function for x and z)	Same precision for x and z?	No. of free parameters	Degrees of freedom	No. of subjects passing nested test 1
$Q_k[x]Q_k[z]$	Yes: linear	Yes	1	5	0
$Q_k[x]Q_n[z]$	Yes: linear	No	2	4	0
$Q_k[w_a(x)]Q_k[w_a(z)]$	Yes: Prelec-2	Yes	3	3	6
$Q_k[w_a(x)]Q_n[w_a(z)]$	Yes: Prelec-2	No	4	2	0
$Q_k[w_a(x)]Q_k[w_c(z)]$	No: Prelec-2	Yes	5	1	11
$Q_k[w_a(x)]Q_n[w_c(z)]$	No: Prelec-2	No	6	-	*3 failed all tests*

Table S2: *Results of nested hypothesis test 1 for experiment 1. There are 20 participants in total.*

Models	Homogeneous? (i.e. same distortion function for x and z)	Same precision for x and z?	No. of free parameters	Degrees of freedom	No. of subjects passing nested test 2
$Q_4[x]Q_4[z]$	Yes: linear	Yes: 4 bits	0	6	0
$Q_k[x]Q_k[z]$	Yes: linear	Yes	1	5	0
$Q_k[x]Q_n[z]$	Yes: linear	No	2	4	0
$Q_4[w_a(x)]Q_4[w_a(z)]$	Yes: Prelec-2	Yes: 4 bits	2	4	3
$Q_k[w_a(x)]Q_k[w_a(z)]$	Yes: Prelec-2	Yes	3	3	3
$Q_k[w_a(x)]Q_n[w_a(z)]$	Yes: Prelec-2	No	4	2	0
$Q_4[w_a(x)]Q_4[w_c(z)]$	No: Prelec-2	Yes: 4 bits	4	2	11
$Q_k[w_a(x)]Q_k[w_c(z)]$	No: Prelec-2	Yes	5	1	1
$Q_k[w_a(x)]Q_n[w_c(z)]$	No: Prelec-2	No	6	-	*2 failed all tests*

Table S3: *Results of nested hypothesis test 2 for experiment 1. There are 20 participants in total. Nested hypothesis test 2 contains three additional 4-bit models.*

Models	Homogeneous? (i.e. same distortion function for x and z)	Same precision for x and z?	No. of free parameters	Degrees of freedom	No. of subjects passing nested test
$Q_4[x]Q_4[z]$	Yes: linear	Yes: 4 bits	0	6	0
$Q_k[x]Q_k[z]$	Yes: linear	Yes	1	5	0
$Q_k[x]Q_n[z]$	Yes: linear	No	2	4	0
$Q_4[w_a(x)]Q_4[w_a(z)]$	Yes: Prelec-2	Yes: 4 bits	2	4	5
$Q_k[w_a(x)]Q_k[w_a(z)]$	Yes: Prelec-2	Yes	3	3	0
$Q_k[w_a(x)]Q_n[w_a(z)]$	Yes: Prelec-2	No	4	2	1
$Q_4[w_a(x)]Q_4[w_c(z)]$	No: Prelec-2	Yes: 4 bits	4	2	10
$Q_k[w_a(x)]Q_k[w_c(z)]$	No: Prelec-2	Yes	5	1	0
$Q_k[w_a(x)]Q_n[w_c(z)]$	No: Prelec-2	No	6	-	0

Table S4: *Results of nested hypothesis test 2 for experiment 2. There are 16 participants in total.*

Models	Homogeneous? (i.e. same distortion function for x, y and z)	Same precision for x, y and z?	No. of free parameters	Degrees of freedom	No. of subjects passing the nested test
$Q_4[x]Q_4[y]Q_4[z]$	Yes: linear	Yes: 4 bits	0	9	1
$Q_k[x]Q_k[y]Q_k[z]$	Yes: linear	Yes	1	8	0
$Q_k[x]Q_m[y]Q_n[z]$	Yes: linear	No	3	6	0
$Q_4[w_a(x)]Q_4[w_a(y)]Q_4[w_a(z)]$	Yes: Prelec-2	Yes: 4 bits	2	7	10
$Q_k[w_a(x)]Q_k[w_a(y)]Q_k[w_a(z)]$	Yes: Prelec-2	Yes	3	6	3
$Q_k[w_a(x)]Q_m[w_a(y)]Q_n[w_a(z)]$	Yes: Prelec-2	No	5	4	4
$Q_4[w_a(x)]Q_4[w_b(y)]Q_4[w_c(z)]$	No: Prelec-2	Yes: 4 bits	6	3	19
$Q_k[w_a(x)]Q_k[w_b(y)]Q_k[w_c(z)]$	No: Prelec-2	Yes	7	2	2
$Q_k[w_a(x)]Q_m[w_b(y)]Q_n[w_c(z)]$	No: Prelec-2	No	9	-	*1 failed all tests*

Table S5: *Results of nested hypothesis test 2 for experiment 3. There are 40 participants in total.*

Appendix 6: Interpretations of Homogeneous and Non-Homogeneous Participants

Mathematically, there is only one true answer to every conjunction probability. However, estimates of conjunctions are subjective, and vary among individuals. In our analysis, we used the quantized Prelec function to allow for these variations. Figure S3A shows the probability distortion curve of a typical homogeneous participant in experiment 1. The dashed diagonal black line denotes the mathematical truth. Since both x and z are distorted similarly, there is only one distortion curve for each participant. Figure S3C shows the distortion curves for a typical non-homogeneous participant, where x and z are distorted differently. Therefore, there is one distortion curve for x and another for z; the curve for the "large" wheel (x) is plotted in red and the "small" wheel (z) in blue. Here, the same objective probability has more than one subjective truth within each individual. The significance is better illustrated in Figure S4. Consider a non-homogeneous participant from our experiment, presented with two separate cases of the 2-wheel task (Figure S4A). In both cases, a wheel with 0.42 probability is presented, but its magnitude is different relative to the other wheel. This participant's probability distortion curves are shown in Figure S4B. The vertical dotted line shows where 0.42 is objectively. This participant processes 0.42 differently in each case, as depicted by the two horizontal arrows in red and blue. Not only does the truth lie in the mind of the beholder, there can be many shades of it internally.

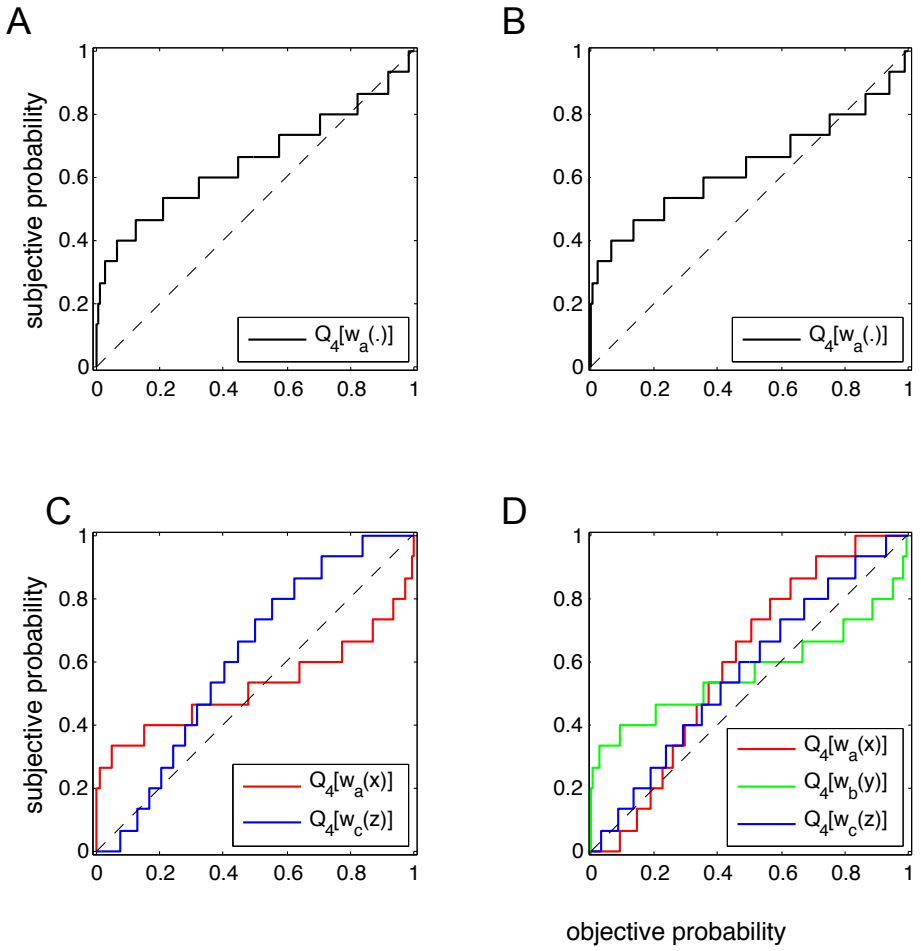


Figure S3: *The probability distortion curves of typical homogeneous and non-homogeneous participants. (A, C) 2-wheel experiment 1. (B, D) 3-wheel experiment 3. (A) w_a represents the quantized Prelec function for the 2 wheels (i.e. x and z); since this is a homogeneous participant, both wheels are distorted in the same manner, and therefore, there is only one common quantized Prelec function. (C) w_a and w_c represent the quantized Prelec functions for wheels x and z respectively. (B) w_a represents the quantized Prelec function for the 3 wheels (i.e. x, y and z); since this is a homogeneous participant, all 3 wheels are distorted in the same manner. (D) w_a, w_b and w_c represent the quantized Prelec functions for wheels x, y and z respectively since this is a non-homogeneous participant.*

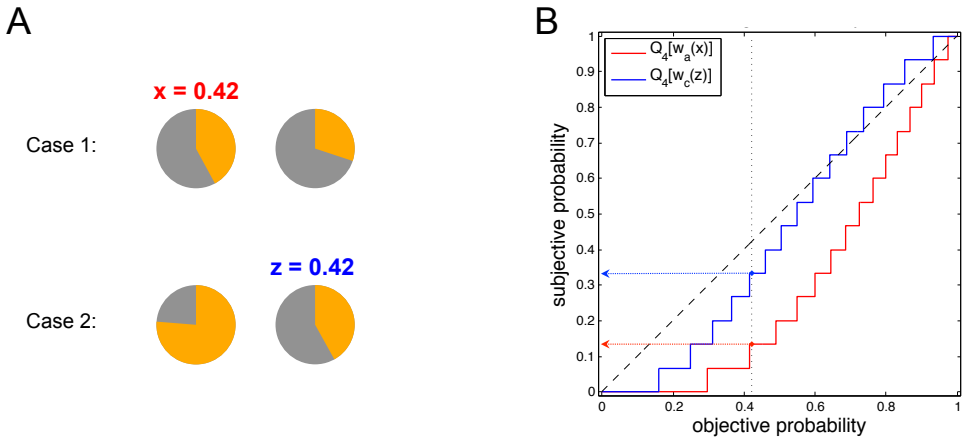


Figure S4: *An objective probability of 0.42 being treated differently depending on its relative magnitude to the other wheel(s). (A) Two separate cases of the 2-wheel task presented to the participant. (B) The corresponding distortion curves for the wheels x and z, with the 2 arrows showing different subjective probabilities corresponding to the objective probability of 0.42.*

Appendix 7: Visual Intuition for Quantized Probability

In order to illustrate the intuition for quantization, we hereby introduce two terminologies: No Noticeable Difference (NND) and Big Noticeable Difference (BND). The first signifies each flat region of the quantized probability distortion curve (any value falling along this flat region is treated indifferently), whereas the second signifies the abrupt step between two flat regions of the curve (the value just after the step is significantly different from the value just before the step). Figure S5A shows the quantized curve for a homogeneous participant from Experiment 3 (i.e. 3-wheel task). The experiment stimuli (i.e. wheels) are plotted as dots. Recall that a homogeneous participant distorts all the stimuli (i.e. wheels x, y and z) using the same quantized curve. The area highlighted by the dashed purple circle contains 3 dots (i.e. stimuli). The second and third dots fall along the flat region, and therefore, both are treated indifferently to one another (No Noticeable Difference). The first dot falls on the flat region that is one step below, and therefore, it is treated differently to the second and third dots (Big Noticeable Difference). Figure S5B shows the corresponding quantized curves and stimuli for a non-homogeneous participant from Experiment 1. In our quantized model, each rise in the graph is an abrupt step rise. We theorize that, in reality, each rise may actually be an S-shaped rise, one that is consistent with the theories of psychometric functions and the Just Noticeable Difference (JND) threshold from Weber's law. Certainly, the very notion of JND implies the existence of NND and BND – meaning, the JND is plausibly sandwiched somewhere in between the NND and the BND.

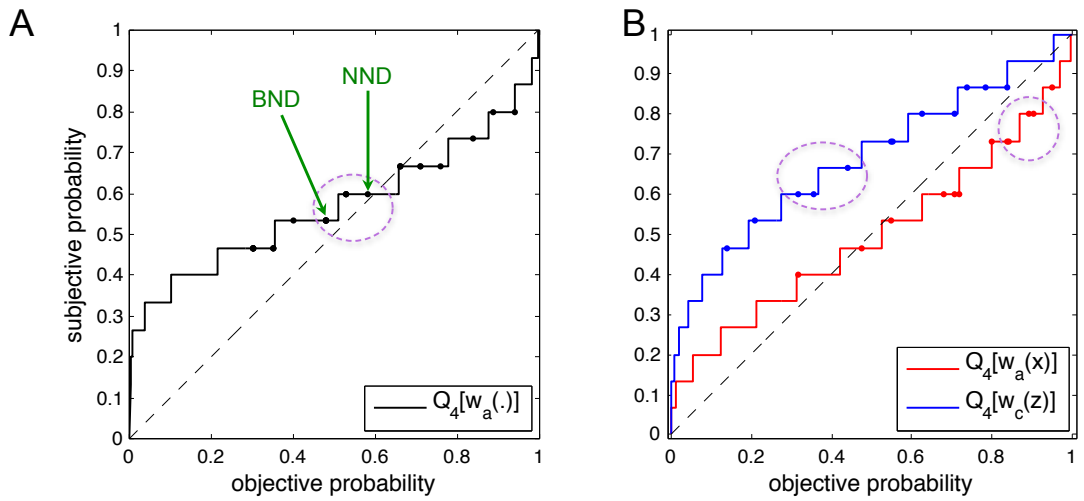


Figure S5: *Dashed purple circles illustrate the notions of No Noticeable Difference (NND) and Big Noticeable Difference (BND). The graphs are quantized probability distortion curves. The stimuli (i.e. wheels of the experiment) are plotted as dots. (A) A homogeneous participant from Experiment 3 (i.e. 3-wheel task). (B) A non-homogeneous participant from Experiment 1 (i.e. 2-wheel task).*

Appendix 8: Summary of All Model Parameters

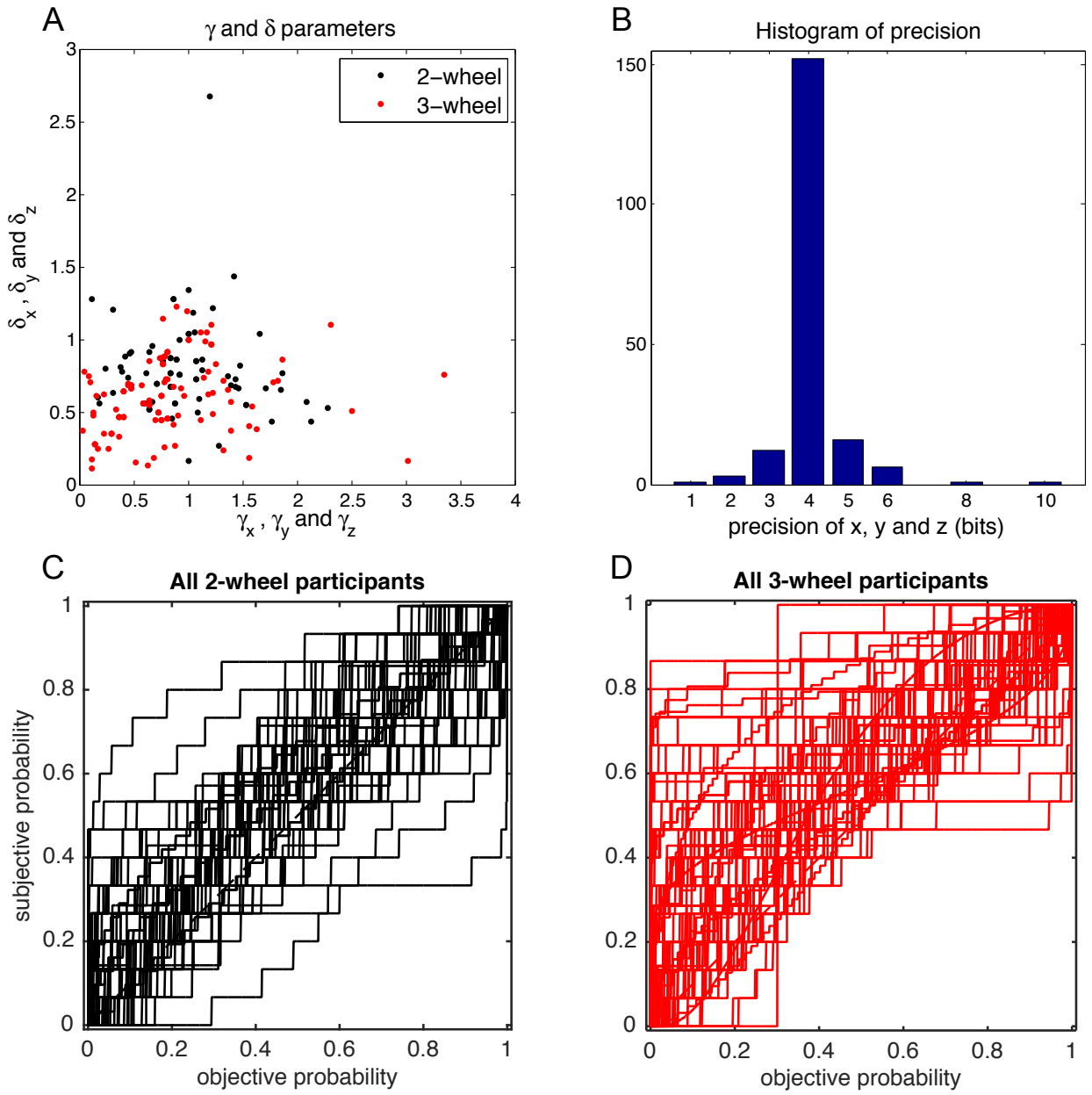


Figure S6: *Summary of the model parameters of all participants. (A) Scatterplot of δ and γ parameters of the quantized probability distortion functions for all participants. There are no obvious clusters or patterns in the parameters values and distortion curves for all participants. (B) Histogram of the precision of all wheels across all participants. The*

precision of all wheels across all participants range from 1 to 10 bits, with a mode of 4 bits. While the locations of the quantized steps (as determined by γ and δ) vary from participant to participant, the number of quantized steps in each wheel is astonishingly similar (i.e. 4 bits = 16 steps), between participants, from wheel to wheel.
(C) Distortion curves for all 2-wheel participants. (D) Distortion curves for all 3-wheel participants.

Appendix 9: A Novel Quantized (Discrete) Exponential Discounting Model

Another commonly used discount function is the continuous exponential discounting model (Laibson, 1997; Chabris et al., 2008):

$$\frac{SV}{A} = \delta^D$$

where SV is the subjective value, A is the objective value, D is the time delay, and δ is the discount rate with $0 < \delta < 1$. We note that in some literature (Green & Myerson, 2004), the exponential discounting model is expressed as:

$$\frac{SV}{A} = e^{-bD}$$

where b is the discount rate parameter. Figure S7 (left) shows an example of a conventional, continuous exponential discounting function. In our work here, we adopted the mathematically equivalent version (Laibson, 1997; Chabris et al., 2008), where:

$$\frac{SV}{A} = e^{-bD} = \left(e^{-b}\right)^D = \left(\frac{1}{e^b}\right)^D = \delta^D$$

Similar to the hyperbolic case, we quantize (Gray & Neuhoff, 1998) this model to produce:

$$Q_n\left[\frac{SV}{A}\right] = Q_n\left[\delta^D\right]$$

where 2^n is the number of steps. Figure S7 (right) shows an example of a 3-bit quantized exponential discounting model (i.e. 3 bits = 2^3 = 8 levels). Similar to the case of a quantized hyperbolic model, the continuous exponential model is simply a quantized model with an infinite number of steps.

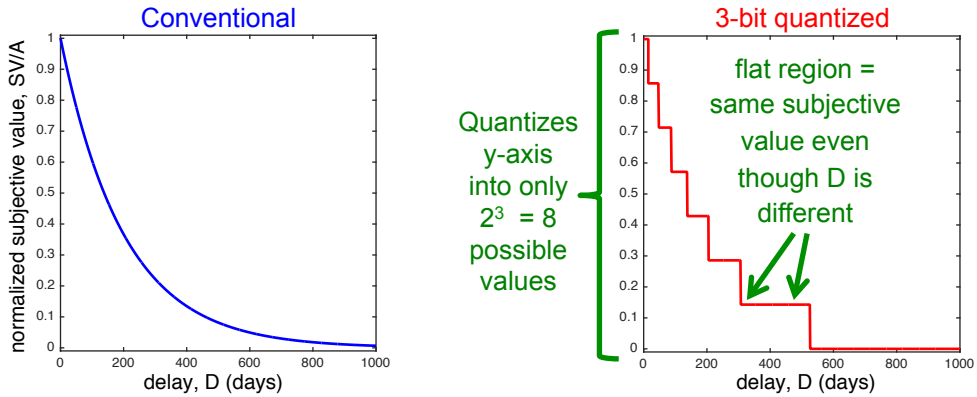


Figure S7: *Left) Conventional, continuous exponential discounting model. Right) 3-bit quantized exponential discounting model.*

Appendix 10: Results for the Quantized Exponential Model

We fit the same experimental data using the same maximum likelihood estimation method to the quantized exponential model. Figure S8 shows the negative log likelihood for a sample participant. Similar to the case of the quantized hyperbolic model, the precisions for the quantized exponential model here range from 1 to 16 bits. The fit for the continuous exponential model is shown in the horizontal dashed red line. As the precision of the quantized exponential model (i.e. blue line) increases from 1 to 5 bits, the fit improves (i.e. value of negative log likelihood decreases). Beyond that, the fit becomes worse (value of negative log likelihood increases) and subsequently flattens off at the level of the continuous model. The best fit occurs at a precision of 5 bits, suggesting that a quantized model may be a better fit than a continuous one, similar to that observed in the hyperbolic case.

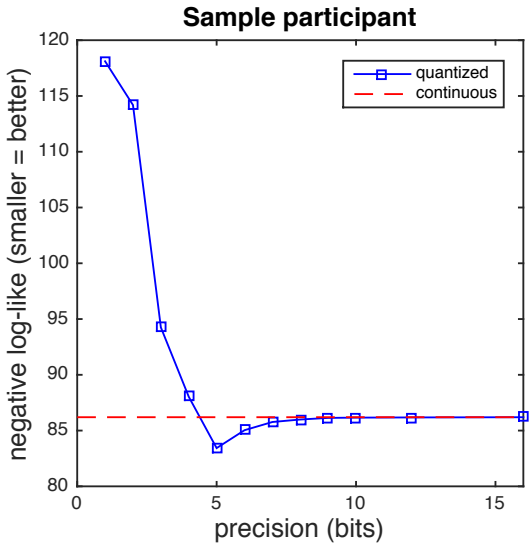


Figure S8: *Negative log likelihood of exponential model fit for one sample participant.*

We found that 8 out of 20 participants were best fit to 5-bit quantized exponential models (i.e. $2^5 = 32$ steps). The histograms of fitted parameters are shown in Figure S9.

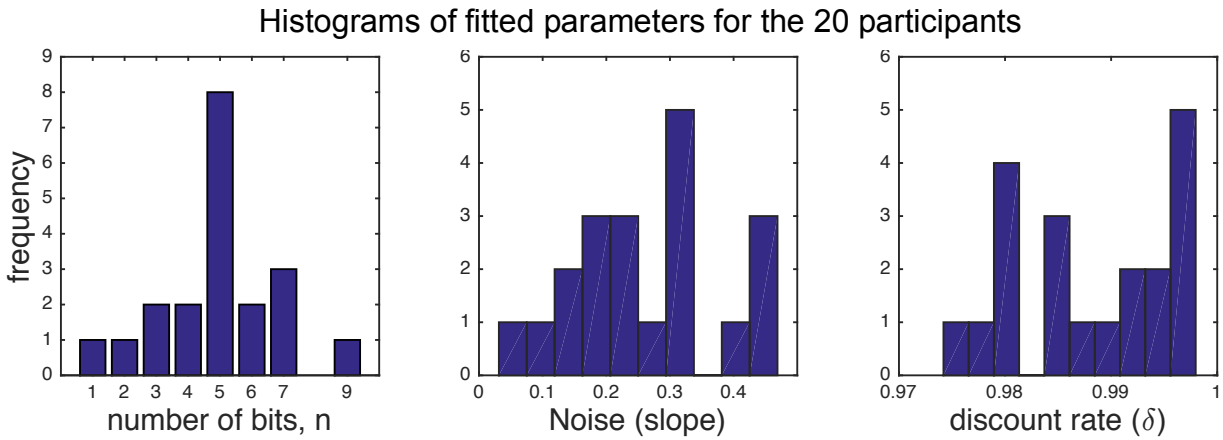


Figure S9: *Histogram of fitted parameters (20 participants) for the quantized exponential model.*

As was the case for the quantized hyperbolic model, we applied a nested hypothesis test to explore whether the second parameter is statistically justifiable or required for the data fitting of each participant:

$$Q_n\left[\delta^D\right] \quad \rightarrow \quad Q_5\left[\delta^D\right]$$

The quantized exponential model on the left has 2 free parameters (i.e. n and δ), whereas the model on the right has only 1 free parameter (i.e. δ) with n being fixed at 5 bits (instead of being a second free parameter). The results from the nested hypothesis test showed that 15 out of 20 participants were best fit to this 1-paramater model (i.e. δ is a free parameter while n is fixed at 5 bits). The 5-bit quantized exponential discounting curves for two representative participants are shown in Figure S10.

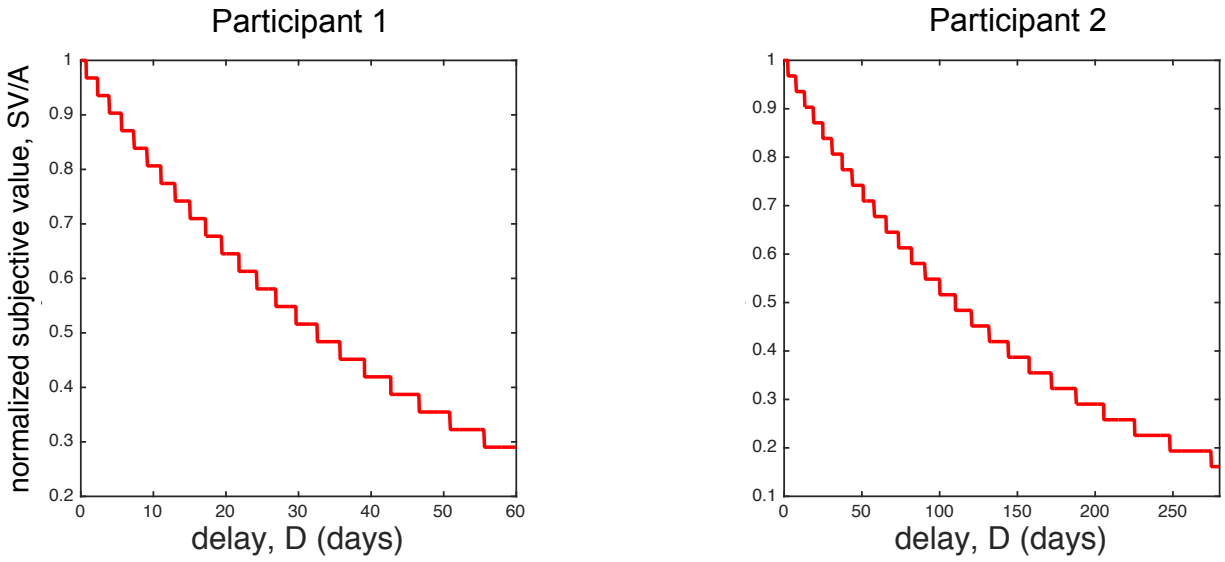


Figure S10: *5-bit quantized exponential discounting curves for two representative participants.*

Appendix 11: Comparing the Quantized Hyperbolic and Quantized Exponential Models

We took the best fit quantized hyperbolic models (i.e. after the nested hypothesis test) and compared it with the best fit quantized exponential models (after the nested hypothesis test) using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (see Busemeyer & Diederich (2014) for an overview of AIC and BIC). These comparisons are plotted in Figure S11 and Figure S12 respectively. Note that, for both the AIC and BIC comparisons, a smaller value represents a better fit. Both AIC and BIC results are in agreement: 13 out of 20 participants were best fit to the quantized exponential model, with the remaining 7 participants best fit to the quantized hyperbolic model.

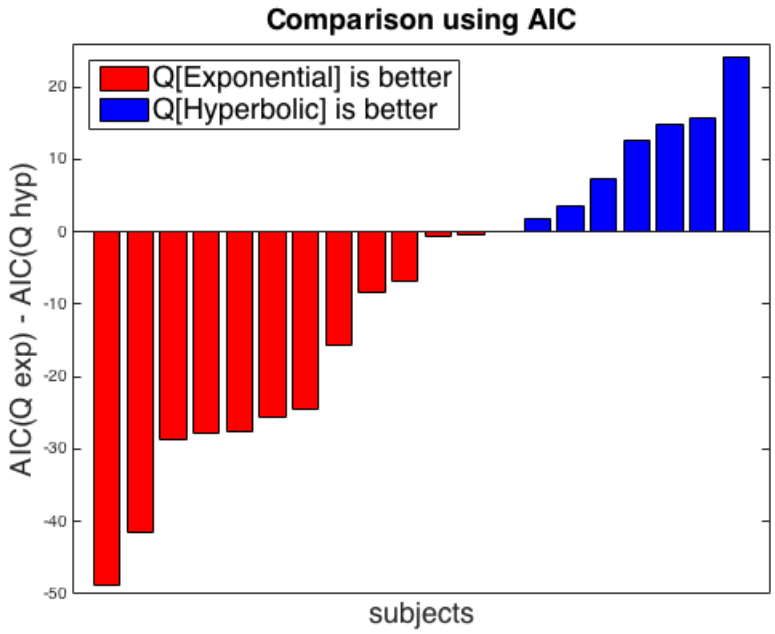


Figure S11: *Comparison using the Akaike Information Criterion (AIC).*

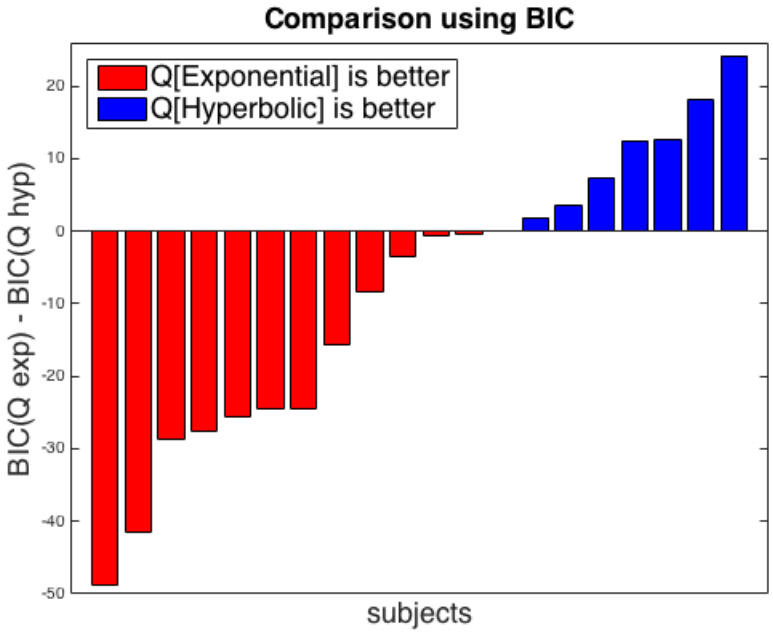


Figure S12: *Comparison using the Bayesian Information Criterion (BIC).*

Following this best-of-the-best AIC/BIC comparison, 15 out of 20 participants have 5-bit precision (see Figure S13).

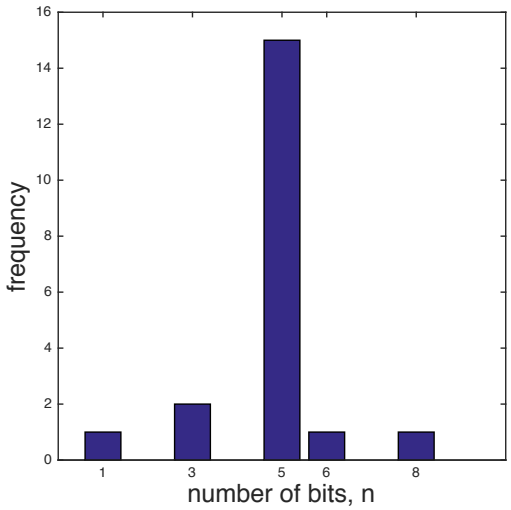


Figure S13: *Histogram of the 20 participants' precisions after the best-of-the-best AIC/BIC comparison.*

A comparison of the quantized exponential and quantized hyperbolic curves of 2 representative participants is shown in Figure S14.

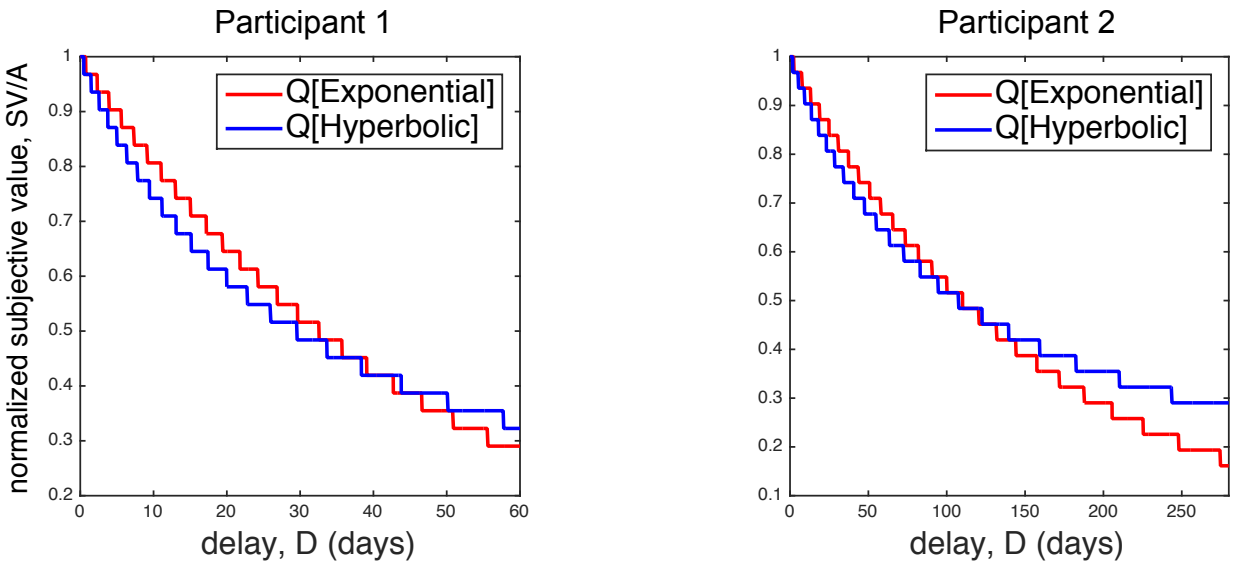


Figure S14: *Comparing the quantized exponential and quantized hyperbolic curves of 2 representative participants.*